A Complementarity Approach to Multistage Stochastic Linear Programs

Dissertation
zur
Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)
vorgelegt der
Mathematisch-naturwissenschaftlichen Fakultät
der
Universität Zürich
von

Simon Siegrist
von
Fahrwangen AG

Promotionskomitee
Prof. Dr. A. D. Barbour (Vorsitz)
Prof. Dr. P. Kall
Prof. Dr. R. Schultz

Zürich 2005

Abstract

The field of Stochastic Programming belongs to the problem class of "Decision-Making under Uncertainty". Applications are widely available in the areas of industrial production and financial planning, among many others. The thesis deals with the approximation of Multistage Stochastic Linear Programs (MSLP) where some model data are assumed to be random and successively realized at time $t = 1, ..., T$ where T is a finite planning horizon. Decisions at time t should be made such that the sum of their immediate costs and the expected recourse costs is minimized, given the previous decisions and the information available up to t. When the number of scenarios is finite, the optimization problem can be formulated as a linear program and may also be solved directly, provided that this number is not too high. Numerical approximation methods are often inevitable, especially if the random data are continuously distributed. There are some methods available for the case $T = 2$ designed for this situation. Unfortunately, they turned out to be impractical to extend to the case $T \geq 3$ because, in this case, the computation of a single recourse function value has almost the same degree of difficulty as determining the optimal objective value of the overall problem.

Since we include the case of continuously distributed data, MSLP is expressed as an infinite linear program which also has an infinite dual form. The optimality gap of a feasible primal-dual pair is expressed as the expectation of a nonnegative random variable, in the thesis called the 'complementarity variable'. Aggregation of constraints and decisions seems to be a natural approach to make MSLP numerically manageable. We analyze particularly models where every optimal solution of a suitably aggregated dual problem is feasible in the original dual problem, leading to lower bounds. After that, based on the aggregated solutions, we propose a way to define recursively a feasible decision policy in the original primal problem by solving a sequence of small linear and quadratic subproblems. Under suitable model assumptions and depending on the aggregation error, the recursive decision policy turns out to be close to the aggregated optimal primal solution. Furthermore, the worst-case behavior of the complementarity variable resulting from the recursive decision policy and the aggregated optimal dual solution is analyzed both in expectation and in an almost sure sense. The latter result is used to prove the finiteness of the proposed refinement algorithm `MSLP-APPROX` which is based on simulated values of the complementarity variable. We also prove that - by successively increasing both the sample size and an accuracy parameter of `MSLP-APPROX` - the (weak) accumulation points of the candidate solutions solve the original problem. In the last part, numerical results are presented in order to illustrate the practical behavior of `MSLP-APPROX`.

Zusammenfassung

Das Gebiet der Stochastischen Programmierung gehört in die Problemklasse der "Entscheidungsfindung unter Unsicherheit". Anwendungen finden sich weitverbreitet in den Feldern der industriellen Produktion und der finanziellen Planung neben vielen anderen. Die Arbeit befasst sich mit der Approximation von 'Multistage Stochastic Linear Programs' (MSLP), wo einige Modelldaten als zufällig vorausgesetzt werden und sich sukzessiv in diskreter Zeit $t = 1, ..., T$ realisieren, wobei T ein endlicher Planungshorizont sei. Entscheidungen zum Zeitpunkt t sollen so gefällt werden, dass die Summe ihrer unmittelbar anfallenden Kosten und den erwarteten Recourse Kosten minimiert wird, gegeben die vorangegangenen Entscheidungen und die Information, welche bis t verfügbar ist. Falls die Anzahl Szenarien endlich ist, dann lässt sich das Optimierungsproblem als Linearprogramm formulieren und auch direkt lösen, sofern diese Anzahl nicht zu gross ist. Numerische Approximationsmethoden sind häufig unumgänglich, insbesondere falls die zufälligen Daten stetig verteilt sind. Es gibt einige Methoden für den Fall $T = 2$, welche auf diese Situation zugeschnitten sind. Leider stellten sich diese als unpraktisch heraus, um sie auf den Fall $T \geq 3$ zu erweitern, weil in diesem Fall die Auswertung eines einzelnen Recourse Funktionswertes nahezu denselben Schwierigkeitsgrad wie die Bestimmung des optimalen Zielfunktionswertes des Gesamtproblems aufweist.

Da wir den Fall von stetig verteilten Daten miteinschliessen, wird MSLP als infinites Linearprogramm formuliert, welches auch eine infinite duale Form besitzt. Die Optimalitätslücke eines zulässigen primal-dual Paares kann als Erwartungswert einer nichtnegativen Zufallsvariablen ausgedrückt werden, in der Arbeit 'Komplementaritätsvariable' genannt. Eine Aggregation von Restriktionen und Entscheidungen scheint ein natürlicher Zugang zu sein, um MSLP numerisch handhabbar zu machen. Wir analysieren vor allem Modelle, bei denen jede optimale Lösung eines geeignet aggregierten Dualproblems zulässig im originalen Dualproblem ist, was auf untere Schranken führt. Danach schlagen wir einen Weg basierend auf den aggregierten Lösungen vor, wie sich rekursiv durch das Lösen einer Folge von kleinen linearen und quadratischen Subproblemen eine zulässige Entscheidungspolitik in der Originalaufgabe definieren lässt. Unter geeigneten Modellannahmen und abhängig vom Aggregationsfehler erweist sich diese Entscheidungspolitik als nahe liegend zu der aggregierten optimalen Primallösung. Ausserdem wird das Worst-Case Verhalten der Komplementaritätsvariable, welche sich aus der rekursiven Entscheidungspolitik und der aggregierten optimalen Duallösung ergibt, sowohl in Erwartung als auch in einem fast sicheren Sinn analysiert. Das letztere Resultat wird verwendet, um die Endlichkeit des vorgeschlagenen Verfeinerungsalgorithmus `MSLP-APPROX` nachzuweisen, welcher auf simulierten Werten der Komplementaritätsvariable basiert. Wir beweisen auch, dass - bei sukzessiver Erhöhung der Stichprobe und eines Genauigkeitsparameters von `MSLP-APPROX` - die (schwachen) Häufungspunkte der Lösungskandidaten das Originalproblem lösen. Um das praktische Verhalten von `MSLP-APPROX` zu veranschaulichen, werden im letzten Teil numerische Resultate präsentiert.

Acknowledgement

I would like to express my sincere thanks and appreciation first of all to Professor Peter Kall for the opportunity to work on this project during a part of my time at the Institute for Operations Research (IOR) of the University of Zurich. I am very grateful for his valuable advice and stimulating suggestions.

Furthermore, I would like to thank Professor Andrew Barbour for his encouragement and his willingness to supervise the thesis at the Institute of Mathematics, Faculty of Science, even though the origin of the topic belongs more to mathematical methods in economics. I am also very grateful to Professor Rüdiger Schultz, University of Duisburg, for reading this dissertation.

I wish to express my gratitude to Professor János Mayer, IOR, for his important comments during many fruitful discussions. My thanks also go to all the other staff members of the IOR, especially to Professor Diethard Klatte, Dr. Peter Fúsek and the Ph.D. students Alexandra Künzi-Bay, Silvia von Bergen, Mike Bielser and Martin Densing for providing valuable support in many respects.

Contents

1 Introduction

1.1 From two to three stages: an illustrative example

Here we illustrate the effect of stochastic constraints on decisions within a multistage model and explain why a natural extension of the simplest known two-stage problem to three stages can become a numerical challenge.

The probably most popular and common example of a (two-stage) stochastic program is given by the classical *newsvendor (or newsboy) problem*. A salesman of newspapers buys $\mathtt{x}_1$ pieces at a price p_1 per newspaper. He can sell them at a price p_2 ($> p_1$). When the demand on his newspapers is η_2 and provided that he tries to sell as many pieces as possible, the return amounts to

$$\mathit{return}(\mathtt{x}_1;\eta_2) = \begin{cases} p_2\mathtt{x}_1 & , \text{ if } \mathtt{x}_1 \le \eta_2 \\ p_2\eta_2 & , \text{ if } \mathtt{x}_1 > \eta_2 \end{cases} \tag{1.1}$$

where we assume first that an unsold newspaper quantity $\mathtt{x}_1 - \eta_2 > 0$ becomes worthless. Suppose now that η_2 is a nonnegative random variable with a continuous distribution function $F(z) := \mathbb{P}[\eta_2 \le z]$ that is assumed to be independent of any decision concerning $\mathtt{x}_1$. One must take into account that $\mathtt{x}_1$ has to be decided *before* η_2 is realized. We set $\mathcal{Q}_2(\mathtt{x}_1;\eta_2) := -\mathit{return}(\mathtt{x}_1;\eta_2)$ as the negative return. The task of determining the best expected (negative) profit results in the problem

$$\mathcal{P}: \quad \min\nolimits_{\mathtt{x}_1 \ge 0} \left\{ p_1\mathtt{x}_1 + \mathbb{E}[\mathcal{Q}_2(\mathtt{x}_1;\eta_2)] \right\} . \tag{1.2}$$

If $\mathbb{E}\eta_2 < +\infty$, then $\mathcal{P}$ is well-defined and the set of minimizers in $\mathcal{P}$ is given by (see e.g. Birge/Louveaux [5])

$$\operatorname{argmin}(\mathcal{P}) = F^{-1}\left(\tfrac{p_2-p_1}{p_2}\right) := \left\{ \mathtt{x}_1 \;\middle|\; F(\mathtt{x}_1) = \tfrac{p_2-p_1}{p_2} \right\} . \tag{1.3}$$

Because $\eta_2 \ge 0$ (a.s.) and $0 < p_1 < p_2$, the restriction '$\mathtt{x}_1 \ge 0$' is not binding in the constrained optimum here, this means it can be dropped. It is worth mentioning that, in general, $\widehat{\mathtt{x}}_1 := \mathbb{E}\eta_2$ is not an optimal solution of $\mathcal{P}$. But $\widehat{\mathtt{x}}_1$ is trivially the solution of the *Expected Value Problem* '$\min_{\mathtt{x}_1 \ge 0}\{p_1\mathtt{x}_1 + \mathcal{Q}_2(\mathtt{x}_1;\mathbb{E}\eta_2)\}$' which is a simplification of the true situation. Looking at (1.1), the negative return is also expressed as

$$\mathcal{Q}_2(\mathtt{x}_1;\eta_2) = -\min\{p_2\mathtt{x}_1, p_2\eta_2\} = \min_{\mathtt{x}_2}\{p_2\mathtt{x}_2 \mid \mathtt{x}_2 \ge -\mathtt{x}_1,\ \mathtt{x}_2 \ge -\eta_2\}. \tag{1.4}$$

Thus, the return (1.1) can also be seen as the result of an optimization. In particular, it is a problem having two parameters $\mathtt{x}_1$ and η_2. The variable (or decision) $\mathtt{x}_2$ is interpreted as the negative selling quantity. First one has to decide on $\mathtt{x}_1$, then one observes a realization of η_2 followed by a recourse action $\mathtt{x}_2$ depending on $\mathtt{x}_1$ and η_2. That is why the model (1.2) is an example for a *two-stage problem*. But, of course, typical two-stage problems are rarely that simple in general. It is very rare in (stochastic) optimization to have an analytical solution as by (1.3). Related models mostly feature a more complex structure with several variables, constraints and random components, which makes any explicit solution formula virtually impossible. We shall now extend (1.2) to a three-stage inventory model in order to demonstrate the latter situation.

With regard to (1.4) one might think that the vendor has really a range to decide on the (negative) selling quantity $\mathtt{x}_2$ at the selling day. For a given $\mathtt{x}_1$ and η_2, even a non-extremal decision $\mathtt{x}_2 > \max\{-\mathtt{x}_1, -\eta_2\}$ is feasible although never optimal in (1.4) since $p_2 > 0$. Thus, there is no physical upper bound for $\mathtt{x}_2$. Maybe the vendor has an interest not to sell as many exemplars as possible, hence he decides for $\max\{-\mathtt{x}_1, -\eta_2\} < \mathtt{x}_2 \leq 0$. Moreover, he could buy even some further newspapers at the market price p_2 (instead of p_1) that would increase his stock. In the latter case one has $\mathtt{x}_2 > 0$. Now suppose that the vendor can sell the remaining (or new) stock $\mathtt{x}_1 + \mathtt{x}_2 > 0$ on a later time but to a different price p_3 and a different demand η_3 which is *unknown* at time 2. If $p_3 > p_2$, then the product "newspapers" is not an adequate example of the issue; but instead, we may think of bottled wine of a special vintage. Assume that the joint distribution of (η_2, η_3) is known and independent of any decision. Generally, $\mathtt{x}_t \geq 0$ means buying whereas $\mathtt{x}_t \leq 0$ means selling the amount of $|\mathtt{x}_t|$ at time t, in both cases at price p_t, $t = 1, 2, 3$. Furthermore, the stock at the end of time 1, 2 and 3 is denoted by $\mathtt{y}_1 = \mathtt{x}_1$, $\mathtt{y}_2 = \mathtt{y}_1 + \mathtt{x}_2$ and $\mathtt{y}_3 = \mathtt{y}_2 + \mathtt{x}_3$. For physical reasons one has to take into account the nonnegativity $\mathtt{y}_t \geq 0$, $(t = 1, 2, 3)$ and $\mathtt{x}_t \geq -\eta_t$ $(t = 2, 3)$; the latter restrictions prevent to sell more than the current demand at time t permits. The problem can now be formulated as a three-stage problem

$$\mathcal{Q}_1 = \min_{\mathtt{x}_1, \mathtt{y}_1} \; p_1 \mathtt{x}_1 + \mathbb{E}[\mathcal{Q}_2(\mathtt{y}_1, \eta_2)] \tag{1.5}$$

$$\begin{aligned} \text{s.t. } \mathtt{y}_1 - \mathtt{x}_1 &= 0 \\ \mathtt{y}_1 &\geq 0, \end{aligned}$$

where

$$\begin{aligned} \mathcal{Q}_2\big(\mathrm{y}_1;\eta_2\big) := \min_{\mathrm{x}_2,\mathrm{y}_2} \quad & p_2\mathrm{x}_2 + \mathbb{E}\big[\mathcal{Q}_3(\mathrm{y}_2,\eta_3) \bigm| \eta_2\big] \\ \text{s.t. } \mathrm{y}_2 - \mathrm{x}_2 &= \mathrm{y}_1 \\ \mathrm{x}_2 &\geq -\eta_2 \\ \mathrm{y}_2 &\geq 0 \end{aligned} \tag{1.6}$$

and

$$\begin{aligned} \mathcal{Q}_3\big(\mathrm{y}_2;\eta_3\big) := \min_{\mathrm{x}_3,\mathrm{y}_3} \quad & p_3\mathrm{x}_3 \\ \text{s.t. } \mathrm{y}_3 - \mathrm{x}_3 &= \mathrm{y}_2 \\ \mathrm{x}_3 &\geq -\eta_3 \\ \mathrm{y}_3 &\geq 0\,. \end{aligned} \tag{1.7}$$

Note that $\mathcal{Q}_3 \equiv 0$ holds whenever $p_3 = 0$. In this case, (1.6) turns out to be identical to (1.4); therefore, (1.5) is the same as the newsvendor problem (1.2). Otherwise, if $p_3 > 0$, it seems to be a hard job to give an analytical solution of (1.5). Even the evaluation of an individual value $\mathcal{Q}_2(\mathrm{y}_1, \eta_2(\omega))$ has at least the same degree of difficulty as determining the optimal objective value of (1.2). Moreover, one has to take the expectation of these values with respect to η_2 so as to minimize the objective in (1.5) over $\mathtt{x}_1 = \mathtt{y}_1 \geq 0$. A similar problem formulation as in (1.5) has been investigated in Lau/Lau [18]. The authors have derived an explicit representation of the objective in (1.5) looking at some different cases for $\mathtt{x}_1$, where it is assumed that η_2 and η_3 are independent and uniformly distributed. The authors refer to numerical methods in order to minimize the objective.

Let us consider the case $0 < p_1 < p_2 < p_3$. Suppose that $(\overline{\mathrm{x}}_1, \overline{\mathrm{y}}_1 = \overline{\mathrm{x}}_1)$ is optimal in (1.5), $(\overline{\mathrm{x}}_2(\eta_2), \overline{\mathrm{y}}_2(\eta_2))$ is optimal in (1.6) given $\mathrm{y}_1 = \overline{\mathrm{y}}_1$, and $(\overline{\mathrm{x}}_3(\eta_2,\eta_3), \overline{\mathrm{y}}_3(\eta_2,\eta_3))$ is optimal in (1.7) given $\mathrm{y}_2 = \overline{\mathrm{y}}_2(\eta_2)$. Of course, the procedure begins with buying a certain amount and ends with selling as much as possible; therefore, it holds $\overline{\mathrm{x}}_1 \geq 0$ and $\overline{\mathrm{x}}_3(\eta_2,\eta_3) \leq 0$ (a.s.). Contrary to the first and last stage, $\overline{\mathrm{x}}_2(\eta_2)$ can take on positive as well as negative values. In the latter case one has to distinguish between $\overline{\mathrm{x}}_2(\eta_2) = -\eta_2$, i.e. selling as much as possible at time 2, and $0 \geq \overline{\mathrm{x}}_2(\eta_2) > -\eta_2$, i.e. covering just a fraction of the demand. Let the demand at the next stage 3 be given as $\eta_3 = \eta_2 + \xi_3$ with ξ_3 being an independent noise. So, when η_2 takes on an unexpectedly high value $\overline{\eta}_2$, the same holds for η_3 in probability. This indicates that possibly $\overline{\mathrm{x}}_2(\overline{\eta}_2) > -\overline{\eta}_2$ (or even $\overline{\mathrm{x}}_2(\overline{\eta}_2) > 0$) holds because one may hope to get rid of the remaining (or new) stock $\overline{\mathrm{y}}_2(\overline{\eta}_2) = \overline{\mathrm{y}}_1 + \overline{\mathrm{x}}_2(\overline{\eta}_2)$

at the highest prize p_3. Contrarily, if η_2 turns out to be unexpectedly small with value $\overline{\eta}_2$, then $\overline{\mathrm{x}}_2(\overline{\eta}_2) = -\overline{\eta}_2$ would be appropriate in order to reduce an oversized stock $\overline{\mathrm{y}}_1$ from the start. In both cases, there is a positive probability to have an undesirable surplus $\overline{\mathrm{y}}_3(\eta_2, \eta_3) > 0$ of no avail at the end. These situations may help to illustrate some questions concerning the structure of optimal decisions affected by stochastic constraints. Of course, finding a near-optimal first-stage decision is the primary goal. But approximating an optimal solution means at the same time determining (or approximating) the optimal subsequent decisions. Optimal decisions at the last stage are given as (pointwise) solutions of linear programs, and therefore, they can always be chosen as vertices of their constraint set. However, decisions at a middle stage depend on a parent decision and are followed by a subsequent operation, and, depending on the output up to that stage, they may or may not be vertices of their feasible set.

In the above example one can show that $\eta_2(\omega) \neq \eta_2(\omega')$ but $\eta_3(\omega) = \eta_3(\omega')$ does in general neither imply $\overline{\mathrm{x}}_3(\eta_2(\omega), \eta_3(\omega)) = \overline{\mathrm{x}}_3(\eta_2(\omega'), \eta_3(\omega'))$ nor $\overline{\mathrm{y}}_3(\eta_2(\omega), \eta_3(\omega)) = \overline{\mathrm{y}}_3(\eta_2(\omega'), \eta_3(\omega'))$. Moreover, both $\overline{\mathrm{x}}_3$ and $\overline{\mathrm{y}}_3$ are not linear combinations of η_2 and η_3. In other words, near-optimal decisions at any stage are nonlinearly influenced by the whole history of the random outcomes. At least, since we assume a finite discrete number of stages, there is a high flexibility in modeling additional constraints on some decisions; the constraints at some stage can basically have a completely different nature than at all the others. Taking our inventory model as an example one might model some capacity constraints on some decisions or a possibility of returning the remaining stock $\mathrm{y}_3(\eta_2, \eta_3)$ at a low prize. All these extensions may have a significant influence on a decision policy.

1.2 Context and basic notations

- Chapter 2 begins with a brief summary of some elementary knowledge about (finite dimensional) linear programs in order to comprehend the basic optimization aspects of the thesis. In Section 2.2 the multistage stochastic linear program (MSLP) is first introduced in the deterministic formulation by using recourse functions. Having three or more stages, the recourse function values can rarely be computed exactly. For this reason, it is probably more convenient to use the infinite linear programming formulation of MSLP, as investigated in Section 2.3 along with the dual MSLP. This helps to separate feasibility- from optimality aspects of decision policies. The connection between the deterministic formulation, the infinite LP- and its infinite dual LP formulation is established in Section 2.4. The optimality gap of a feasible primal-dual pair will be expressed as the expectation of a nonnegative random variable, in the thesis called the *complementarity variable*.

- In Chapter 3, after having briefly discussed some well known approaches to stochastic recourse models (Sections 3.1-3.3), the aggregation principles according to Wright [30] are analyzed and simultaneously extended to the case of some randomness in the technology matrices (Section 3.4). Section 3.5 proposes a way to define recursively a primal feasible policy based on the aggregated solutions. The policy is determined by solving a sequence of small linear- and quadratic subproblems.

- The validation of the recursive policy resulting from Section 3.5 is studied in Chapter 4. For this purpose, it is necessary to analyze first the sensitivity of a unique selection of nonunique LP solutions (Section 4.1). In Section 4.2 we make suitable assumptions in such a way that, depending on the aggregation error, the recursive policy turns out to be close to the aggregated optimal primal solution. Furthermore, the worst-case behavior of the complementarity variable resulting from the recursive policy in combination with the aggregated optimal dual solution is analyzed both in expectation and in an almost sure sense.

- The main subject of Chapter 5 is a proposed operational algorithm for the approximation of MSLP, named `MSLP-APPROX`. The method is based on simulated values of the complementarity variable in order to make suitable refinements of a rectangle structure. The stability results of Chapter 4 are used to prove the finiteness of the method. After that we make a probabilistic inference on the complementarity variable with respect to its expected smallness at the stopping time of `MSLP-APPROX`. This leads to an infinite extension, named `MSLP-SOLVE`, where we prove that - by successively increasing the sample size and the accuracy parameter of `MSLP-APPROX` - the (weak) accumulation points of the candidate solutions solve the original problem.

- Numerical results are presented in Chapter 6 in order to illustrate the practical behavior of `MSLP-APPROX`. We provide some computational results for the inventory model (1.5) - extended to T stages, together with a few nontrivial 2, 3, 4 and 5-stage examples with a total of up to 24 continuously distributed random elements.

- At the end we recapitulate some facets of the proposed approximation scheme in connection with a discussion about possible extensions.

Many symbols in the thesis represent matrix- or vector valued random variables denoted by ξ or A, b, c, x, etc. A distinction to deterministic variables/parameters/data should be deduced from the context. The Euclidean norm in $\mathbb{R}^n$ is denoted by $|\cdot|$ in each dimension, where $|D|$, $D \in \mathbb{R}^{m\times n}$, means the induced Euclidean matrix norm of D. We suppose that all random variables are defined on one and the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$. If $x : \Omega \to \mathbb{R}^n$ is measurable with respect to $\mathcal{A}$, then the measurability is abbreviated by $x \sim \mathcal{A}$. The *expectation* of x is denoted by $\mathbb{E}x := \int_\Omega x(\omega) d\mathbb{P}(\omega)$ - if it exists - and ess sup$|x|$ denotes the *essential supremum* of $|x|$ with respect to $\mathbb{P}$. The random variables are frequently restricted to a *sub-σ-algebra* $\mathcal{G}$ of $\mathcal{A}$. The *L^p-spaces* on $\mathcal{G}$ are denoted by

$$\begin{aligned} L^p(\mathcal{G}; \mathbb{R}^n) &:= \big\{x : \Omega \to \mathbb{R}^n \mid x \sim \mathcal{G},\ \mathbb{E}|x|^p < \infty\big\}\ (p \geq 1), \text{ and} \\ L^\infty(\mathcal{G}; \mathbb{R}^n) &:= \big\{x : \Omega \to \mathbb{R}^n \mid x \sim \mathcal{G},\ \text{ess sup}|x| < \infty\big\}, \end{aligned}$$

respectively; (since Ω and $\mathbb{P}$ are fixed we do not mention them here). If $\xi : \Omega \longrightarrow \mathbb{R}^l$, $\xi \sim \mathcal{A}$, then $\sigma(\xi) := \sigma\big(\{\xi^{-1}(B) \,|\, B \subset \mathbb{R}^l \text{ Borel measurable}\}\big)$ is the smallest σ-field of $\mathcal{A}$ with respect to which ξ is measurable. The smallest σ-field in $\mathcal{A}$ containing $\mathcal{G}_1, \mathcal{G}_2 \subset \mathcal{A}$ is denoted by $\sigma(\mathcal{G}_1, \mathcal{G}_2)$, whereas $\sigma(\mathcal{G}_1, \xi) := \sigma\big(\mathcal{G}_1, \sigma(\xi)\big)$. $\mathbb{E}[\xi \,|\, \mathcal{G}]$ is the *conditional expectation of* $\xi \sim \mathcal{A}$ *with respect to the σ-field* $\mathcal{G}$. An elementary knowledge of some properties of conditional expectations is required for the understanding ("take out what is known", "exchange the order of expectations", etc.). The *support* (in the image space) of ξ is defined by

$$\text{supp}\,\{\xi\} := \bigcap_{S \in \mathcal{S}} S \quad \big(\subset \mathbb{R}^l\big)$$

where $\mathcal{S} := \big\{S \subset \mathbb{R}^l \,|\, S \text{ closed},\ \mathbb{P}[\xi \in S] = 1\big\}$. If x, y are $\mathbb{R}^n$-valued, then $x \leq y$ (a.s.) means $x_i \leq y_i$ (a.s.) $(i = 1, ..., n)$. We let $\mathbb{R}^n_+ := \{z \in \mathbb{R}^n \,|\, z \geq 0\}$. The *interior* of $B \subset \mathbb{R}^m$ is denoted by int$B$ and the *convex hull* of B is conv$B$. Let X be an arbitrary set and $f : X \to \mathbb{R}$, $C \subset X$. When an optimization problem

$$(\mathcal{P}) : \quad \text{Minimize}_x\big\{f(x) \;\big|\; x \in C\big\}$$

is labeled, $\inf(\mathcal{P}) \in [-\infty, +\infty]$ means its optimal value where $\inf(\mathcal{P}) := +\infty$ if $C = \emptyset$. We also write $\min(\mathcal{P}) \in (-\infty, +\infty)$ when a minimum exists, and argmin$(\mathcal{P})$ $\big(\subset C\big)$ denotes the set of minimizers of $\mathcal{P}$. Each $x \in C$ is said to be a *feasible solution* of $\mathcal{P}$, and in addition, if $f(x) \leq \inf(\mathcal{P}) + \varepsilon$, then x is called an ε-minimizer of $\mathcal{P}$.

2 Multistage Stochastic Linear Programming

2.1 Basics of Linear Programming

This section gives a short survey on some fundamental LP theory which is relevant in the thesis. A (finite dimensional) LP in standard form reads as

$$(\mathcal{P}): \quad \min_{x \in \mathbb{R}^n} \left\{ c^\top x \;\middle|\; Ax = b,\ x \geq 0 \right\} \tag{2.1}$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$. In the case that the constraints look more generally as $Ax = b,\ Cx \geq d$, with $C \in \mathbb{R}^{l \times n}$, $d \in \mathbb{R}^l$, one can add a slack vector z leading to $Cx - z = d,\ z \geq 0$. Furthermore, a reformulation of x into $x = x^+ - x^-$ with $x^+, x^- \geq 0$ leads to the appropriate form (2.1). As the standard dual of the primal problem $(\mathcal{P})$ we refer to

$$(\mathcal{D}): \quad \max_{u \in \mathbb{R}^m, s \in \mathbb{R}^n} \left\{ b^\top u \;\middle|\; A^\top u + s = c,\ s \geq 0 \right\} \tag{2.2}$$

where u is unconstrained in sign. Note that $A\mathbb{R}^n_+$ is the set of all b with a feasible $(\mathcal{P})$, whereas $A^\top \mathbb{R}^m + \mathbb{R}^n_+$ is the set of all c with a feasible $(\mathcal{D})$. The following proposition belongs to the standard results of linear programming. For further details we refer to *Dantzig* [7] or *Kall/Wallace* [14].

Proposition 2.1 (cf. [7], [14])

a) *(Weak duality)*
If x is feasible in $(\mathcal{P})$ and (u, s) is feasible in $(\mathcal{D})$, then their objectives differ by

$$c^\top x - b^\top u = c^\top x - (Ax)^\top u = (c - A^\top u)^\top x = s^\top x \geq 0\ .$$

b) *(Strong duality)*
$(\mathcal{P})$ is solvable if and only if $(\mathcal{P})$ and $(\mathcal{D})$ are feasible. In this case there is at least one feasible x in $(\mathcal{P})$ and one feasible (u, s) in $(\mathcal{D})$ satisfying $s^\top x = 0$. Hence a) implies that this pair is optimal and $\min(\mathcal{P}) = \max(\mathcal{D})$.

c) *Let $(\mathcal{D})$ be feasible and let $\gamma(b) = \min(\mathcal{P})$ denote the optimal objective value of $(\mathcal{P})$ depending on the parameter $b \in A\mathbb{R}^n_+$. Then $\gamma : A\mathbb{R}^n_+ \longrightarrow \mathbb{R}$ is convex and piecewise linear.*

d) Let $\mathcal{B}(b) := \{x \geq 0 \mid Ax = b\}$ denote the feasible set of (2.1) depending on the parameter $b \in A\mathbb{R}^n_+$. Then $\mathcal{B}(b)$ is bounded if and only if $\{z \geq 0 \mid Az = 0\} = \{0\}$. In this case, $\mathcal{B}(b)$ is bounded for all b and there is a constant $L = L(A)$ so that $|x| \leq L|b|$, $\forall x \in \mathcal{B}(b)$, $\forall b \in \mathbb{R}^m$.

e) There exists a constant $L = L(A) > 0$ such that $\left|\binom{u}{s}\right| \leq L|c|$ for all optimal solutions (u, s) of $(\mathcal{D})$, $\forall b \in int(A\mathbb{R}^n_+)$, $\forall c \in (A^\top\mathbb{R}^m + \mathbb{R}^n_+)$.

Note: $int(A\mathbb{R}^n_+)$ is nonempty if and only if A has full row rank.

Remarks 2.2

1) For solving a linear program on a PC platform exactly with say, $n \leq 10^5$, we refer to standard LP solvers as `GAMS/CPLEX` or `GAMS/MINOS`. Both are based on a simplex method. They are available with the General Algebraic Modeling System `GAMS`, see Brooke et al. [6]. The solver `MINOS` is also suitable for nonlinear programs (NLP). Other solvers can sometimes be more efficient, based for example on decomposition methods by exploiting a special structure of A, see also later Section 3.1. Since the present work deals mainly with the subsequent processing of LP solutions derived from discretized models, these well-known LP algorithms will not be further outlined.

2) We remark that in c), $\gamma(\cdot)$ is linear on finitely many (and implicitly given) convex polyhedral cones. Suppose that $\tilde{b}$ is a random vector in $A\mathbb{R}^n_+$ with existing expectation where the support of $\tilde{b}$ intersects several of these cones. Then the expected optimal value $\mathbb{E}\gamma(\tilde{b})$ in general cannot be computed exactly. At least, since $A\mathbb{R}^n_+$ is convex, Jensen's inequality yields a deterministic lower bound by $\gamma(\mathbb{E}\tilde{b}) \leq \mathbb{E}\gamma(\tilde{b})$. For deterministic upper bounds we refer to Section 3.3.

3) Some theoretical difficulties can arise under perturbation of the matrix A. To explain this by an example, take the problem

$$\gamma(t) := \min_{x \in \mathbb{R}} \{x \mid tx = t,\ x \geq 0\}.$$

It obviously holds $\gamma(0) = 0$, whereas $\gamma(t) = 1$, $t \neq 0$. Thus, if $\overline{\xi}$ takes on the value 0 with probability 1, and if $\xi^{(k)}$ converges in distribution to $\overline{\xi}$ with $\xi^{(k)}(\omega) \neq 0$ for almost all ω, then $0 = \mathbb{E}\gamma(\overline{\xi}) \neq \lim_{k\to\infty} \mathbb{E}\gamma(\xi^{(k)}) = 1$. In general, therefore, one cannot expect any stability of the expected LP objective under perturbation of the distribution when A is not fixed.

At least, see Kall [13] Thm. 1, the LP objective γ depending on all data A, b and c is a Borel-measurable extended real-valued function $\gamma : \mathbb{R}^{m\times n+m+n} \to [-\infty, \infty]$.

2.2 The deterministic formulation of MSLP

Given a finite time set $\{1, ..., T\}$, a multistage stochastic linear program (MSLP) is usually interpreted as the problem of determining

$$\begin{aligned} \mathcal{Q}_1 \quad := \quad & \inf_{\mathrm{y}_1\in\mathbb{R}^{n_1}} c_1^\top \mathrm{y}_1 + \mathbb{E}\left[\mathcal{Q}_2(\mathrm{y}_1; \vec{\eta}_2)\right] \\ & \text{s.t. } A_1\mathrm{y}_1 = b_1 \\ & \qquad \mathrm{y}_1 \geq 0 \end{aligned} \tag{2.3}$$

where for $t = 2, ..., T$, $\mathrm{y}_{t-1} \in \mathbb{R}^{n_{t-1}}$ and $\omega \in \Omega$, the *recourse function* is pointwise given by

$$\begin{aligned} \mathcal{Q}_t\big(\mathrm{y}_{t-1}; \vec{\eta}_t(\omega)\big) \quad := \quad & \inf_{\mathrm{y}_t\in\mathbb{R}^{n_t}} c_t(\omega)^\top \mathrm{y}_t + \mathbb{E}\big[\mathcal{Q}_{t+1}(\mathrm{y}_t; \vec{\eta}_{t+1}) \,\big|\, \vec{\eta}_t = \vec{\eta}_t(\omega)\big] \\ & \text{s.t. } A_t(\omega)\mathrm{y}_t = b_t(\omega) - B_t(\omega)\mathrm{y}_{t-1} \\ & \qquad \mathrm{y}_t \geq 0, \end{aligned} \tag{2.4}$$

and

- $\mathcal{Q}_{T+1} \equiv 0$,
- $\eta_1 := (A_1, b_1, c_1)$ is known at time 1 where $A_1 \in \mathbb{R}^{m_1\times n_1}$, $b_1 \in \mathbb{R}^{m_1}$ and $c_1 \in \mathbb{R}^{n_1}$,
- $\eta_t := (A_t, B_t, b_t, c_t)$ has its realization at time $t \in \{2, ..., T\}$ where

  $$A_t : \Omega \to \mathbb{R}^{m_t\times n_t}, \;\; B_t : \Omega \to \mathbb{R}^{m_t\times n_{t-1}}, \;\; b_t : \Omega \to \mathbb{R}^{m_t}, \; c_t : \Omega \to \mathbb{R}^{n_t}$$

  are random valued on a joint probability space $(\Omega, \mathcal{A}, \mathbb{P})$,
- $\vec{\eta}_t := (\eta_1, ..., \eta_t)$ is the information available by the time $t = 1, .., T$.

Note that, for some $\omega \in \Omega$, the problems (2.3) or (2.4) can be infeasible (therefore the optimal value is defined as $+\infty$) or unbounded ($= -\infty$). For the purpose of well-defined integrals, for any mapping $Z : \Omega \to [-\infty, +\infty]$ and $\eta \sim \mathcal{A}$, we define $\mathbb{E}[Z \,|\, \eta = \eta(\omega)]$ as

$$\inf_{\widehat{Z}\in L^1(\mathcal{A};\mathbb{R}),\, \widehat{Z}\geq Z\,(\text{a.s.})} \mathbb{E}[\widehat{Z} \,|\eta = \eta(\omega)] \quad \big(\in [-\infty, +\infty)\big), \tag{2.5}$$

if $\exists \widehat{Z} \in L^1(\mathcal{A};\mathbb{R})$, $\widehat{Z} \geq Z$ (a.s.), or as $+\infty$ else. Note that the convention is consistent insofar as it coincides with the usual definition whenever $Z \in L^1(\mathcal{A};\mathbb{R})$.

The aim of (2.3) and (2.4) is as clear as simple: at each stage we shall minimize the immediate cost $c_t(\omega)^\top \mathrm{y}_t$ plus the expected optimal costs arising at a later time given all the information by the time t. Moreover, no further costs arise at any time $t > T$. The dimensions $(m_t \times n_t)$ are supposed to be rather small, varying between say, (1×2) and (10×20), with a number of stages T being 2 up to say 10 or more. In the majority of applications, only a few components of $\vec{\eta}_T : \Omega \to \mathbb{R}^{\sum_{t=1}^{T}(m_t n_t + m_t n_{t-1} + m_t + n_t)}$ are non-constant, that is to say random. However, we presume at least one component of η_t to be non-constant, $t = 2, ..., T$. In the inventory model in Section 1.1 as an example, see (1.6)-(1.7), only the right-hand side vectors b_2 and b_3 are actually random.

We remark that (2.3) is a nonlinear problem in $\mathbb{R}^{n_1}$. The advantage of this representation is obviously given by its meaningful interpretation. But for any $\mathrm{y}_1 \in \mathbb{R}^{n_1}$, in general, we are not able to evaluate its exact objective $c_1^\top \mathrm{y}_1 + \mathbb{E}\left[\mathcal{Q}_2(\mathrm{y}_1;\vec{\eta}_2)\right]$. Moreover, if $T \geq 3$, then even the evaluation of an integrand value $\mathcal{Q}_2(\mathrm{y}_1;\vec{\eta}_2(\omega))$ is almost as complicated as determining the optimal objective $\mathcal{Q}_1$. That is why (2.3) is not treated as an application of nonlinear (convex) programming because the pointwise evaluation of the objectives and subgradients is not cheap at all. In general, (2.3) seems to be far away from giving an exact solution, also in the case $T = 2$. There are some special settings as

- $\{\eta_t\}_{t=2}^T$ being stochastically independent ("interstage-independence"),
- or at least, the process $(\eta_t)_{t=2}^T$ having the Markov property.

In these cases, it can be easily seen by a backward induction argument that one may write $\mathcal{Q}_t\big(\mathrm{y}_{t-1};\eta_t\big)$ instead of $\mathcal{Q}_t\big(\mathrm{y}_{t-1};\vec{\eta}_t\big)$. Moreover, interstage-independence implies $\mathbb{E}\big[\mathcal{Q}_{t+1}(\mathrm{y}_t;\vec{\eta}_{t+1}) \;\big|\; \vec{\eta}_t\big] = \mathbb{E}\big[\mathcal{Q}_{t+1}(\mathrm{y}_t;\eta_{t+1}) \;\big|\; \vec{\eta}_t\big] = \mathbb{E}[\mathcal{Q}_{t+1}(\mathrm{y}_t;\eta_{t+1})]$. But these special cases make the problem not significantly easier with regard to numerical aspects. In fact, interstage-independence is always given in the two-stage case which is itself not trivial at all, whereas the Markov property always holds for $T \leq 3$.

2.3 The infinite LP formulation and its LP dual

The purpose here is to derive a primal-dual pair of infinite linear programs associated with the deterministic form (2.3). We also give some rather general assumptions in order to get strong duality. As a main consequence in the next section, both-sided bounds on the recourse functions (2.4) result and this is particularly so for the overall problem (2.3) as well.

To avoid possible confusion we remark that the following duality framework concerns a relaxation of the constraints in (2.3) and (2.4) rather than a relaxation of the so called ‘nonanticipativity’ of decisions, cf. Rockafellar/Wets [27]. In that paper together with Rockafellar/Wets [26] and e.g. Olsen [21], [22], similar problems as in (2.3) are analyzed with respect to equivalent or dual formulations. The authors look for existence of optimal solutions together with sufficient and necessary optimality conditions. However, the assumptions made in these papers are hardly conformable with each other. For example, in [22] the whole matrix structure $(B_t, A_t)_{t=2,...,T}$ is presumed to be fixed, whereas in [26] the subject is reduced to $T = 2$ with convex constraints.

Because we restrict our studies to the linear case, dual statements can be derived from the theory of *Infinite Linear Programming* (ILP) without using Lagrange-Duality (including the case of inequality constraints) or Kuhn-Tucker Theory. In this respect, an intuitive background is provided by the very basic LP duality in $\mathbb{R}^n$, see Section 2.1. From our viewpoint, MSLP is a brilliant application for the theory of ILP rather than of Nonlinear Programming. However, sometimes we must refer to the book of Anderson/Nash [2] because linear programming in infinite dimensional spaces is expected to be ‘slightly’ different than in $\mathbb{R}^n$. A review can also be found in Anderson [1].

Recall that $\vec{\eta}_t = (\eta_1, ..., \eta_t)$ is the data information available by the time $t = 1, ..., T$. Let $\mathcal{F}_t := \sigma(\vec{\eta}_t) \subset \mathcal{A}$ be the sub-σ-algebra of $\mathcal{A}$ induced by $\vec{\eta}_t$. Hence we are concerned with a filtration $\mathcal{F} := (\mathcal{F}_t)_{t=1,...,T}$ in $(\Omega, \mathcal{A}, \mathbb{P})$,

$$\{\emptyset, \Omega\} = \mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_T \subset \mathcal{A}\ .$$

In particular, we generally assume

Model Assumptions 2.3
Suppose that $A_1 \in \mathbb{R}^{m_1 \times n_1}$, $b_1 \in \mathbb{R}^{m_1}$, $c_1 \in \mathbb{R}^{n_1}$, *and for* $t = 2, ..., T$,

- $A_t \in L^\infty(\mathcal{F}_t; \mathbb{R}^{m_t \times n_t})$, $B_t \in L^\infty(\mathcal{F}_t; \mathbb{R}^{m_t \times n_{t-1}})$,
- $b_t \in L^2(\mathcal{F}_t; \mathbb{R}^{m_t})$, $c_t \in L^2(\mathcal{F}_t; \mathbb{R}^{n_t})$.

Instead of (2.3), we shall understand MSLP alternatively as

$$\texttt{MSLP-}\mathcal{P}: \qquad \text{Minimize}_x \ \mathbb{E}\left[\sum_{t=1}^T c_t^\top x_t\right]$$

$$\text{subject to} \begin{cases} A_1 x_1 &= b_1 & & \\ B_t x_{t-1} + A_t x_t &= b_t & \text{(a.s.)} & (t = 2, ..., T) \\ x_t &\geq 0 & \text{(a.s.)} & (t = 1, ..., T) \\ x_t &\in L^2(\mathcal{F}_t; \mathbb{R}^{n_t}) & & (t = 1, ..., T) \end{cases} \tag{2.6}$$

where $L^2(\mathcal{F}_1; \mathbb{R}^{n_1}) \cong \mathbb{R}^{n_1}$. The above formulation is given as one single minimization problem; but both the number of linear constraints and variables is infinite whenever $\mathcal{F}_T$ is not finitely generated. Contrary to (2.3), a main advantage of this infinite representation is given by a better distinction between 'feasible' and 'optimal' policies. First we shall derive its LP dual and thereafter, in Section 2.4, a link to (2.3) is given by both-sided bounds on the recourse functions (2.4), in particular, bounds on the unknown optimal objective value $\mathcal{Q}_1$ in (2.3).

Notations 2.4
We occasionally make use of the following notations:

$$\mathbf{A} := \begin{pmatrix} A_1 & & & \\ B_2 & A_2 & & \\ & \ddots & \ddots & \\ & & B_T & A_T \end{pmatrix}, \ \mathbf{b} := \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_T \end{pmatrix}, \ \mathbf{c} := \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_T \end{pmatrix},$$

and with $m := (m_t)_{t=1}^T$, $n := (n_t)_{t=1}^T$, $\mathcal{F} = (\mathcal{F}_t)_{t=1}^T$, *we use*

$$\begin{aligned} L^2(\mathcal{F}; \mathbb{R}^m) &:= L^2(\mathcal{F}_1; \mathbb{R}^{m_1}) \times \cdots \times L^2(\mathcal{F}_T; \mathbb{R}^{m_T}), \\ L^2(\mathcal{F}; \mathbb{R}^n) &:= L^2(\mathcal{F}_1; \mathbb{R}^{n_1}) \times \cdots \times L^2(\mathcal{F}_T; \mathbb{R}^{n_T}), \end{aligned}$$

both equipped with the scalar product $\mathbb{E}[(\cdot)^\top(\cdot)]$.

The space $L^2(\mathcal{F};\mathbb{R}^n)$ is a linear subspace of $L^2(\mathcal{A};\mathbb{R}^n) := L^2(\mathcal{A};\mathbb{R}^{n_1}) \times \cdots \times L^2(\mathcal{A};\mathbb{R}^{n_T})$ where the so called *nonanticipativity* of decisions would fail, i.e., $x_t \sim \mathcal{F}_t$ would be relaxed to $x_t \sim \mathcal{A}$, $(t = 1, ..., T)$. However, our approach is not based on this relaxation, and therefore, decisions will always be adapted to $L^2(\mathcal{F};\mathbb{R}^n)$.

According to Model Assumptions 2.3 it holds $\mathbf{b} \in L^2(\mathcal{F};\mathbb{R}^m)$ and $\mathbf{c} \in L^2(\mathcal{F};\mathbb{R}^n)$. Moreover, $\mathbf{A}$ induces a linear and continuous mapping $\mathbf{A} : L^2(\mathcal{F};\mathbb{R}^n) \to L^2(\mathcal{F};\mathbb{R}^m)$ from one Hilbert space into the other. A decision (or say policy) $x = (x_1^\top, ..., x_T^\top)^\top \in L^2(\mathcal{F};\mathbb{R}^n)$ is feasible in MSLP-$\mathcal{P}$ if and only if $\mathbf{A}x = \mathbf{b}$, $x \geq 0$, (a.s.). Thus one may think of MSLP-$\mathcal{P}$ as a special linear program in Hilbert spaces given by

$$\texttt{MSLP-}\mathcal{P}: \quad \text{Minimize}_{x\in L^2(\mathcal{F};\mathbb{R}^n)}\big\{\mathbb{E}[\mathbf{c}^\top x] \;\big|\; \mathbf{A}x = \mathbf{b},\; x \geq 0\;,\; \text{(a.s.)}\big\}. \tag{2.7}$$

Because Hilbert spaces are self-dual, the dual multipliers can be chosen in L^2 as well. Following a general LP duality theory in Anderson/Nash [2], the dual of (2.7) reads as

$$\texttt{MSLP-}\mathcal{D}: \quad \text{Maximize}_{u\in L^2(\mathcal{F};\mathbb{R}^m),\, s\in L^2(\mathcal{F};\mathbb{R}^n)}\big\{\mathbb{E}[\mathbf{b}^\top u] \;\big|\; \mathbf{A}^*u + s = \mathbf{c},\; s \geq 0,\, \text{(a.s.)}\big\} \tag{2.8}$$

in which $\mathbf{A}^* : L^2(\mathcal{F};\mathbb{R}^m) \to L^2(\mathcal{F};\mathbb{R}^n)$ is the adjoint mapping of $\mathbf{A}$. It is given as

$$\mathbf{A}^*u = \begin{pmatrix} A_1^\top u_1 + \mathbb{E}[B_2^\top u_2 \,|\, \mathcal{F}_1] \\ A_2^\top u_2 + \mathbb{E}[B_3^\top u_3 \,|\, \mathcal{F}_2] \\ \vdots \\ A_{T-1}^\top u_{T-1} + \mathbb{E}[B_T^\top u_T \,|\, \mathcal{F}_{T-1}] \\ A_T^\top u_T \end{pmatrix}, \quad u = (u_1^\top, ..., u_T^\top)^\top \in L^2(\mathcal{F};\mathbb{R}^m)\;, \tag{2.9}$$

noting that $\mathbb{E}[B_2^\top u_2 \,|\, \mathcal{F}_1] = \mathbb{E}[B_2^\top u_2]$. The reason why $\mathbf{A}^*$ is the adjoint mapping is justified by

Lemma 2.5
It holds $\mathbf{A}^*u \in L^2(\mathcal{F};\mathbb{R}^n)$, $\forall u \in L^2(\mathcal{F};\mathbb{R}^m)$, and

$$\mathbb{E}[(\mathbf{A}x)^\top u] = \mathbb{E}[(\mathbf{A}^*u)^\top x],\; \forall x \in L^2(\mathcal{F};\mathbb{R}^n), \forall u \in L^2(\mathcal{F};\mathbb{R}^m)\;.$$

Proof. The first part follows immediately from the definition of $\mathbf{A}^*$. Furthermore, for any $x \in L^2(\mathcal{F};\mathbb{R}^n)$ and $u \in L^2(\mathcal{F};\mathbb{R}^m)$ holds

$$\begin{aligned}
\mathbb{E}[(\mathbf{A}x)^\top u] &= \mathbb{E}[(\mathbf{A}^\top u)^\top x] = \mathbb{E}\left[\sum_{t=1}^{T-1}(A_t^\top u_t + B_{t+1}^\top u_{t+1})^\top x_t + (A_T^\top u_T)^\top x_T\right] \\
&= \sum_{t=1}^{T-1}\mathbb{E}\left[(A_t^\top u_t + B_{t+1}^\top u_{t+1})^\top x_t\right] + \mathbb{E}\left[(A_T^\top u_T)^\top x_T\right] \\
&= \sum_{t=1}^{T-1}\mathbb{E}\Big[\mathbb{E}\left[(A_t^\top u_t + B_{t+1}^\top u_{t+1})^\top x_t \middle| \mathcal{F}_t\right]\Big] + \mathbb{E}\left[(A_T^\top u_T)^\top x_T\right] \\
&= \sum_{t=1}^{T-1}\mathbb{E}\Big[\mathbb{E}\left[A_t^\top u_t + B_{t+1}^\top u_{t+1} \middle| \mathcal{F}_t\right]^\top x_t\Big] + \mathbb{E}\left[(A_T^\top u_T)^\top x_T\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T-1}\left(A_t^\top u_t + \mathbb{E}\left[B_{t+1}^\top u_{t+1} \middle| \mathcal{F}_t\right]\right)^\top x_t + (A_T^\top u_T)^\top x_T\right] \\
&= \mathbb{E}[(\mathbf{A}^* u)^\top x]\ . \qquad \square
\end{aligned}$$

In terms of geometry, $\mathbf{A}^*u$ is the (Euclidean) projection of $\mathbf{A}^\top u \in L^2(\mathcal{A};\mathbb{R}^n)$ onto $L^2(\mathcal{F};\mathbb{R}^n)$. We remark that the primal-dual pair (2.7)-(2.8) is composed similarly to (2.1)-(2.2) of Section 2.1. The main difference is in the generalization of the transposed matrix to the adjoint mapping. In accordance with the primal original (2.6) we state (2.8) more explicitly as

$$\texttt{MSLP-}\mathcal{D}: \qquad \text{Maximize}_{u,s}\ \mathbb{E}\left[\textstyle\sum_{t=1}^{T} b_t^\top u_t\right]$$

$$\text{subject to}\ \left\{\begin{array}{rcll}
A_t^\top u_t + \mathbb{E}[B_{t+1}^\top u_{t+1}\,|\,\mathcal{F}_t] + s_t &=& c_t\ \text{(a.s.)} & (t=1,...,T-1) \\
A_T^\top u_T + s_T &=& c_T\ \text{(a.s.)} & \\
s_t &\geq& 0\ \ \text{(a.s.)} & (t=1,...,T) \\
u_t &\in& L^2(\mathcal{F}_t;\mathbb{R}^{m_t}) & (t=1,...,T) \\
s_t &\in& L^2(\mathcal{F}_t;\mathbb{R}^{n_t}) & (t=1,...,T)
\end{array}\right. . \tag{2.10}$$

This infinite dual version of `MSLP` has also been obtained in Wright [30] but rather via Lagrange duality (without a discussion on strong duality). The basic ideas of that paper will be discussed and extended in Section 3.4 and 3.5.

Weak and strong duality

If $x = (x_1, ..., x_T)$ and $(u, s) = \big((u_1, s_1), ..., (u_T, s_T)\big)$ are feasible in $\texttt{MSLP-}\mathcal{P}$ and $\texttt{MSLP-}\mathcal{D}$, respectively, then their objective values differ by

$$\begin{aligned} \mathbb{E}[\mathbf{c}^\top x] - \mathbb{E}[\mathbf{b}^\top u] &= \mathbb{E}[\mathbf{c}^\top x] - \mathbb{E}[(\mathbf{A}x)^\top u] = \mathbb{E}[\mathbf{c}^\top x] - \mathbb{E}[(\mathbf{A}^* u)^\top x] \\ &= \mathbb{E}[(\mathbf{c} - \mathbf{A}^* u)^\top x] = \mathbb{E}[\underbrace{s^\top x}_{\geq 0 \text{ (a.s.)}}] \geq 0 \;, \end{aligned} \tag{2.11}$$

or in detail,

$$\mathbb{E}\left[\sum_{t=1}^{T} c_t^\top x_t\right] - \mathbb{E}\left[\sum_{t=1}^{T} b_t^\top u_t\right] = \mathbb{E}\left[\sum_{t=1}^{T} \underbrace{s_t^\top x_t}_{\geq 0 \text{ (a.s.)}}\right] \geq 0 \;. \tag{2.12}$$

It follows that the weak duality

$$\sup(\texttt{MSLP-}\mathcal{D}) \leq \inf(\texttt{MSLP-}\mathcal{P}) \tag{2.13}$$

holds. We will call $\sum_{t=1}^{T} s_t^\top x_t = s^\top x : \Omega \to [0, +\infty)$ the *complementarity variable of* x *and* (u, s). Some rather strong (and rarely verifiable) conditions on the model must be made to ensure strong duality as in

Proposition 2.6 (cf. Anderson/Nash [2], Thm. 10 and Thm. 22)
Let $-\infty < \inf(\texttt{MSLP-}\mathcal{P}) < +\infty$*, and in addition, let the set*

$$\left\{(\mathbf{A}x, \mathbb{E}[\mathbf{c}^\top x]) \;\middle|\; x \in L^2(\mathcal{F}; \mathbb{R}^n),\; x \geq 0\right\} \subset L^2(\mathcal{F}; \mathbb{R}^m) \times \mathbb{R} \tag{2.14}$$

be closed (in the norm $\|\cdot\| := \{\mathbb{E}[(\cdot)^\top(\cdot)]\}^{\frac{1}{2}} + |\cdot|$ *). Then* $\texttt{MSLP-}\mathcal{P}$ *is solvable. Moreover it holds* $\sup(\texttt{MSLP-}\mathcal{D}) = \min(\texttt{MSLP-}\mathcal{P})$*, in particular,* $\texttt{MSLP-}\mathcal{D}$ *is feasible.*

Remarks 2.7

1) *Under the closure condition (2.14) and because of the weak duality (2.13), the properties '*$-\infty < \inf(\texttt{MSLP-}\mathcal{P}) < +\infty$*' and '*$\texttt{MSLP-}\mathcal{P}$ *&* $\texttt{MSLP-}\mathcal{D}$ *are feasible' turn out to be equivalent. On the other hand, if the closure condition fails, then even* $-\infty < \sup(\texttt{MSLP-}\mathcal{D}) < \inf(\texttt{MSLP-}\mathcal{P}) < +\infty$ *is sometimes possible, cf. Anderson/Nash [2]. In the finite discrete case, see also later Section 3.1, the closure condition is always satisfied due to the LP theory in terms of Section 2.1.*

2) *It is worth mentioning that some necessary assumptions needed for an "operational approach" to an infinite problem are expected to be very different (and stronger!) than the closure condition (2.14) providing the absence of a positive duality gap. At the end in Proposition 5.6 we derive constructively a feasible dual sequence* $(\widehat{u}^{(r)}, \widehat{s}^{(r)})_{r\geq 1}$ *and a feasible primal sequence* $(\overline{x}^{(r)})_{r\geq 1}$ *satisfying* $\lim_{r\to\infty} \mathbb{E}[\widehat{s}^{(r)\top}\overline{x}^{(r)}] = 0$, *and every weak accumulation point solves the problem. It is a question of convenience whether the strong duality result of Proposition 2.6 is preferable, going without many technical details, or whether a kind of "operational duality" is adequate by using many more assumptions on the model.*

2.4 Bounds on the deterministic problem

We turn back to the deterministic formulation (2.3) of MSLP by using the recourse functions (2.4).

Convention: The notation $x = (x_1, ..., x_T)$ is reserved for a policy in $L^2(\mathcal{F}; \mathbb{R}^n)$, whereas $\mathtt{y}_t \in \mathbb{R}^{n_t}$ is a real-valued decision at stage t, $(t = 1, ..., T)$. Analogously, $u = (u_1, ..., u_T)$ and $s = (s_1, ..., s_T)$ are reserved for elements in $L^2(\mathcal{F}; \mathbb{R}^m)$ and $L^2(\mathcal{F}; \mathbb{R}^n)$, respectively, whereas $\mathtt{v}_t \in \mathbb{R}^{m_t}$ and $\mathtt{r}_t \in \mathbb{R}^{n_t}$, $(t = 1, ..., T)$. Furthermore, $\mathcal{Q}_1(x_0, \vec{\eta}_1) := \mathcal{Q}_1(\mathtt{y}_0, \vec{\eta}_1) := \mathcal{Q}_1$ means the optimal value of (2.3) and we let $B_1 := 0$.

Lemma 2.8
Let $x = (x_1, ..., x_T)$ *and* $(u, s) = \big((u_1, s_1), ..., (u_T, s_T)\big)$ *be feasible in* `MSLP-`$\mathcal{P}$ *and* `MSLP-`$\mathcal{D}$*, respectively. Then for* $t = 1, ..., T$ *it holds*

$$\mathcal{Q}_t(x_{t-1}, \vec{\eta}_t) \leq \mathbb{E}\left[\sum_{r=t}^{T} c_r^\top x_r \;\middle|\; \vec{\eta}_t\right] \text{ (a.s.)}, \tag{2.15}$$

whereas

$$\mathcal{Q}_t(\mathtt{y}_{t-1}, \vec{\eta}_t) \geq \mathbb{E}\left[\sum_{r=t}^{T} b_r^\top u_r \;\middle|\; \vec{\eta}_t\right] - \mathtt{y}_{t-1}^\top B_t^\top u_t \text{ (a.s.)},\ \forall \mathtt{y}_{t-1} \in \mathbb{R}^{n_{t-1}}\ . \tag{2.16}$$

Proof. We prove (2.15) and (2.16) by a backward induction argument on $t = T, T-1, ..., 1$. Because $x_t(\omega)$ is a feasible continuation of $x_{t-1}(\omega)$ for almost all ω, $t = 2, ..., T$, we first obtain $\mathcal{Q}_T(x_{T-1}, \vec{\eta}_T) \leq c_T^\top x_T = \mathbb{E}[c_T^\top x_T \,|\, \vec{\eta}_T]$, (a.s.), and if (2.15) holds for $t+1$, then

$$\begin{aligned}
\mathcal{Q}_t(x_{t-1},\vec{\eta}_t) &= \inf_{\mathtt{y}_t\geq 0,\, A_t\mathtt{y}_t=b_t-B_tx_{t-1}} c_t^\top \mathtt{y}_t + \mathbb{E}\big[\mathcal{Q}_{t+1}(\mathtt{y}_t;\vec{\eta}_{t+1}) \bigm| \vec{\eta}_t\big] \\
&\leq c_t^\top x_t + \mathbb{E}\big[\mathcal{Q}_{t+1}(x_t;\vec{\eta}_{t+1}) \bigm| \vec{\eta}_t\big] \\
&\leq c_t^\top x_t + \mathbb{E}\left[\mathbb{E}\big[\textstyle\sum_{r=t+1}^{T} c_r^\top x_r \bigm| \vec{\eta}_{t+1}\big] \,\middle|\, \vec{\eta}_t\right] \\
&= \mathbb{E}\left[\textstyle\sum_{r=t}^{T} c_r^\top x_r \,\middle|\, \vec{\eta}_t\right], \text{ (a.s.)}.
\end{aligned}$$

This proves (2.15). On the other hand, for any $\mathtt{y}_{T-1}\in\mathbb{R}^{n_{T-1}}$, we have the pointwise defined LP (in the sense of Section 2.1)

$$\begin{aligned}
\mathcal{Q}_T\big(\mathtt{y}_{T-1};\vec{\eta}_T\big) \;=\; & \min\nolimits_{\mathtt{y}_T\in\mathbb{R}^{n_T}} c_T^\top \mathtt{y}_T \\
& \text{s.t. } A_T\mathtt{y}_T = b_T - B_T\mathtt{y}_{T-1} \\
& \qquad \mathtt{y}_T \geq 0 \;,
\end{aligned}$$

with its LP-dual

$$\begin{aligned}
& \max\nolimits_{\mathtt{v}_T\in\mathbb{R}^{m_T},\mathtt{r}_T\in\mathbb{R}^{n_T}} \big(b_T - B_T\mathtt{y}_{T-1}\big)^\top \mathtt{v}_T \\
& \text{s.t. } A_T^\top \mathtt{v}_T + \mathtt{r}_T = c_T \\
& \qquad \mathtt{r}_T \geq 0 \;.
\end{aligned}$$

Note that (u_T, s_T) is dual feasible almost surely (cf. (2.10)). Thus

$$\mathcal{Q}_T\big(\mathtt{y}_{T-1};\vec{\eta}_T\big) \geq \big(b_T - B_T\mathtt{y}_{T-1}\big)^\top u_T = \mathbb{E}[b_T^\top u_T \,|\, \vec{\eta}_T] - \mathtt{y}_{T-1}^\top (B_T^\top u_T) \;\; \text{(a.s.)}.$$

So let us assume that (2.16) holds for $t+1$. Then one has for any $\mathtt{y}_{t-1}\in\mathbb{R}^{n_{t-1}}$,

$$\begin{aligned}
\mathcal{Q}_t(\mathtt{y}_{t-1},\vec{\eta}_t) &= \inf_{\mathtt{y}_t\geq 0,\, A_t\mathtt{y}_t=b_t-B_t\mathtt{y}_{t-1}} c_t^\top \mathtt{y}_t + \mathbb{E}\big[\mathcal{Q}_{t+1}(\mathtt{y}_t;\vec{\eta}_{t+1}) \bigm| \vec{\eta}_t\big] \\
&\geq \inf_{\mathtt{y}_t\geq 0,\, A_t\mathtt{y}_t=b_t-B_t\mathtt{y}_{t-1}} c_t^\top \mathtt{y}_t + \mathbb{E}\left[\mathbb{E}\big[\textstyle\sum_{r=t+1}^{T} b_r^\top u_r \bigm| \vec{\eta}_{t+1}\big] - \mathtt{y}_t^\top B_{t+1}^\top u_{t+1} \,\middle|\, \vec{\eta}_t\right] \\
&= \mathbb{E}\left[\textstyle\sum_{r=t+1}^{T} b_r^\top u_r \,\middle|\, \vec{\eta}_t\right] + \min_{\mathtt{y}_t\geq 0,\, A_t\mathtt{y}_t=b_t-B_t\mathtt{y}_{t-1}} \big(c_t - \mathbb{E}\left[B_{t+1}^\top u_{t+1} \,\middle|\, \vec{\eta}_t\right]\big)^\top \mathtt{y}_t.
\end{aligned}$$

Again, the problem on the right is a pointwise defined LP which has its dual (and again in the sense of Section 2.1)

$$\begin{aligned}
& \max\nolimits_{\mathtt{v}_t\in\mathbb{R}^{m_t},\mathtt{r}_t\in\mathbb{R}^{n_t}} \big(b_t - B_t\mathtt{y}_{t-1}\big)^\top \mathtt{v}_t \\
& \text{s.t. } A_t^\top \mathtt{v}_t + \mathtt{r}_t = c_t - \mathbb{E}\left[B_{t+1}^\top u_{t+1} \,\middle|\, \vec{\eta}_t\right] \\
& \qquad \mathtt{r}_t \geq 0 \;.
\end{aligned}$$

Because (u_t, s_t) is a.s.-feasible in this problem (cf. (2.10)), it follows that

$$\begin{aligned}\mathcal{Q}_t(\mathtt{y_{t-1}}, \vec{\eta}_t) &\geq \mathbb{E}\left[\textstyle\sum_{r=t+1}^{T} b_r^\top u_r \;\middle|\; \vec{\eta}_t\right] + \left(b_t - B_t \mathtt{y}_{t-1}\right)^\top u_t \\ &= \mathbb{E}\left[\textstyle\sum_{r=t}^{T} b_r^\top u_r \;\middle|\; \vec{\eta}_t\right] - \mathtt{y}_{t-1}^\top B_t^\top u_t \ , \ \text{(a.s.)}.\end{aligned}$$

This completes the proof of (2.16). □

Theorem 2.9
Let $x = (x_1, ..., x_T)$ and $(u, s) = \big((u_1, s_1), ..., (u_T, s_T)\big)$ be feasible in `MSLP-`$\mathcal{P}$ and `MSLP-`$\mathcal{D}$, respectively. Then for $t = 1, ..., T$ it holds

$$\mathbb{E}\left[\sum_{r=t}^{T} b_r^\top u_r \;\middle|\; \vec{\eta}_t\right] - x_{t-1}^\top B_t^\top u_t \leq \mathcal{Q}_t(x_{t-1}, \vec{\eta}_t) \leq \mathbb{E}\left[\sum_{r=t}^{T} c_r^\top x_r \;\middle|\; \vec{\eta}_t\right] \ , \ \text{(a.s.)}, \tag{2.17}$$

where the right and left bounds differ by

$$\mathbb{E}\left[\sum_{r=t}^{T} s_r^\top x_r \;\middle|\; \vec{\eta}_t\right] \ \text{(a.s.)}. \tag{2.18}$$

In particular, for $t = 1$, let $\varepsilon := \mathbb{E}\left[\sum_{r=1}^{T} s_r^\top x_r \;\middle|\; \vec{\eta}_1\right] = \mathbb{E}[s^\top x]$. Then one has

$$\mathbb{E}[\mathbf{b}^\top u] \leq \sup(\texttt{MSLP-}\mathcal{D}) \leq \mathcal{Q}_1 \leq \inf(\texttt{MSLP-}\mathcal{P}) \leq \mathbb{E}[\mathbf{c}^\top x]$$

and the bounds differ by ε; therefore, $x_1 \in L^2(\mathcal{F}_1; \mathbb{R}^{n_1}) \cong \mathbb{R}^{n_1}$ is an ε-minimizer in (2.3). Moreover, if the closure condition (2.14) of Proposition 2.6 is satisfied, let therefore $\overline{x} = (\overline{x}_1, ..., \overline{x}_T)$ be a solution of `MSLP-`$\mathcal{P}$, then $\overline{x}_1$ is a minimizer in (2.3).

Proof. The first part (2.17) follows immediately from Lemma 2.8 where $\mathtt{y}_t \in \mathbb{R}^{n_t}$ is replaced by $x_t(\omega)$, $t = 1, ..., T$. The representation (2.18) is simply based on the feasibility of x and (u, s) as well as on the 'taking out what is known' property of conditional expectations because for any $t \in \{1, ..., T\}$ it holds

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{r=t}^{T} c_r^\top x_r \,\middle|\, \vec{\eta}_t\right] - \left(\mathbb{E}\left[\sum_{r=t}^{T} b_r^\top u_r \,\middle|\, \vec{\eta}_t\right] - x_{t-1}^\top B_t^\top u_t\right) \\
= \;& \mathbb{E}\left[\sum_{r=t}^{T} c_r^\top x_r \,\middle|\, \vec{\eta}_t\right] - \mathbb{E}\left[\sum_{r=t}^{T} (B_r x_{r-1} + A_r x_r)^\top u_r \,\middle|\, \vec{\eta}_t\right] + x_{t-1}^\top B_t^\top u_t \\
= \;& \mathbb{E}\left[\sum_{r=t}^{T} c_r^\top x_r \,\middle|\, \vec{\eta}_t\right] - \mathbb{E}\left[\sum_{r=t+1}^{T} (B_r^\top u_r)^\top x_{r-1} \,\middle|\, \vec{\eta}_t\right] - \mathbb{E}\left[\sum_{r=t}^{T} (A_r^\top u_r)^\top x_r \,\middle|\, \vec{\eta}_t\right] \\
= \;& \mathbb{E}\left[\sum_{r=t}^{T} c_r^\top x_r \,\middle|\, \vec{\eta}_t\right] - \mathbb{E}\left[\sum_{r=t+1}^{T} \left(\mathbb{E}[B_r^\top u_r \,|\, \vec{\eta}_{r-1}]\right)^\top x_{r-1} \,\middle|\, \vec{\eta}_t\right] \\
& - \mathbb{E}\left[\sum_{r=t}^{T} (A_r^\top u_r)^\top x_r \,\middle|\, \vec{\eta}_t\right] \\
= \;& \mathbb{E}\left[\sum_{r=t}^{T} c_r^\top x_r \,\middle|\, \vec{\eta}_t\right] - \mathbb{E}\left[\sum_{r=t}^{T-1} \left(\mathbb{E}[B_{r+1}^\top u_{r+1} \,|\, \vec{\eta}_{r}]\right)^\top x_{r} \,\middle|\, \vec{\eta}_t\right] \\
& - \mathbb{E}\left[\sum_{r=t}^{T} (A_r^\top u_r)^\top x_r \,\middle|\, \vec{\eta}_t\right] \\
= \;& \mathbb{E}\left[\sum_{r=t}^{T-1} \left(c_r - \mathbb{E}[B_{r+1}^\top u_{r+1} \,|\, \vec{\eta}_r] - A_r^\top u_r\right)^\top x_r \,\middle|\, \vec{\eta}_t\right] \\
& + \mathbb{E}\left[\left(c_T - A_T^\top u_T\right)^\top x_T \,\middle|\, \vec{\eta}_t\right] \\
= \;& \mathbb{E}\left[\sum_{r=t}^{T-1} s_r^\top x_r \;\middle|\; \vec{\eta}_t\right] + \mathbb{E}\left[s_T^\top x_T \;\middle|\; \vec{\eta}_t\right] = \mathbb{E}\left[\sum_{r=t}^{T} s_r^\top x_r \;\middle|\; \vec{\eta}_t\right] \;, \text{(a.s.)}.
\end{aligned}
$$

The last part is an immediate consequence of Proposition 2.6. □

If the sequences $(x^{(r)})_{r\geq 1}$ and $(u^{(r)}, s^{(r)})_{r\geq 1}$ consist of feasible elements of MSLP-$\mathcal{P}$ and MSLP-$\mathcal{D}$, respectively, where $\lim_{r\to\infty} \mathbb{E}[s^{(r)\top} x^{(r)}] = 0$, then this also means by the above Theorem that the objective values in (2.3) of the first-stage solutions $(x_1^{(r)})_{r\geq 1}$ converge to the finite infimum $\mathcal{Q}_1$. For this reason we will focus on the problems MSLP-$\mathcal{P}$ and MSLP-$\mathcal{D}$, whereas with a few exceptions the deterministic formulation (2.3) will be omitted in the present work.

3 Approximation

At first, we briefly discuss the benefits and drawbacks of some well known approaches to stochastic recourse models with the aim of

- solving the problem exactly (Section 3.1),
- approximating the distribution (Section 3.2),
- designing both-sided deterministic bounds (Section 3.3).

After that in Section 3.4, the lower bounding principles according to Wright [30] are analyzed and extended to some further model classes. Our idea to obtain (statistical) upper bounds by recursions is presented in Section 3.5.

3.1 The finite discrete case

Recall the basic model assumptions on page 12, but suppose that the filtration $\mathcal{F} = (\mathcal{F}_t)_{t=1,\dots,T}$ in $(\Omega, \mathcal{A}, \mathbb{P})$,

$$\{\emptyset, \Omega\} = \mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_T \subset \mathcal{A} ,$$

is finitely generated, that is to say, $\mathcal{F}_T$ is a finitely generated sub-σ-algebra of $\mathcal{A}$. The integrability conditions on the data are therefore superfluous and a nodal syntax is probably more adequate in this finite discrete case (see also Fig. 3.1).

Notations 3.1
A (rooted) T-stage tree consists of a node set $\mathcal{N} = \bigcup_{t=1}^{T} \mathcal{N}_t$ where $\mathcal{N}_t \cap \mathcal{N}_r = \emptyset$, $t \neq r$; $\mathcal{N}_1 = \{1\}$ is the root and $\mathcal{N}_t$ the node set at stage t. Furthermore, each $n \in \mathcal{N}_t$ features a unique immediate ancestor node $p_n \in \mathcal{N}_{t-1}$, whereas each $m \in \mathcal{N}_{t-1}$ possesses a nonempty set of immediate successor nodes ('children'), i.e.,

$$C_m := \{n \in \mathcal{N}_t \mid p_n = m\} \neq \emptyset \ \ , \ \forall m \in \mathcal{N}_{t-1} \ (t = 2, ..., T) \ .$$

If the tree $\mathcal{N} = (\mathcal{N}_t)_{t=1,\dots,T}$ is associated with $\mathcal{F} = (\mathcal{F}_t)_{t=1,\dots,T}$, then each node $n \in \mathcal{N}_t$ belongs to a set $\Omega^{[n]} \in \mathcal{F}_t$ so that $\mathcal{F}_t = \sigma\big((\Omega^{[n]})_{n \in \mathcal{N}_t}\big)$ where $\Omega^{[1]} = \Omega$ and, in particular, each $\Omega^{[m]}$ is a disjoint union of its immediate successor sets, i.e.

$$\Omega^{[m]} = \bigcup_{n \in C_m} \Omega^{[n]} \ , \ \ \Omega^{[n]} \cap \Omega^{[n']} = \emptyset, \ n, n' \in C_m, n \neq n',$$

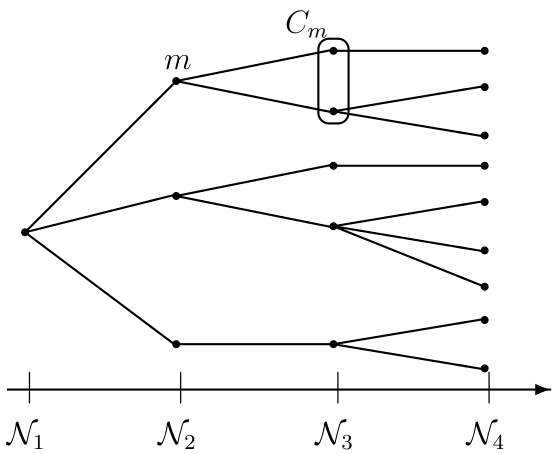

Figure 3.1: *A scenario tree with $T = 4$. Here, the labeled node $m \in \mathcal{N}_2$ has two immediate successor nodes.*

for all $m \in \mathcal{N}_{t-1}$ $(t = 2, ..., T)$. Assume that for each $n \in \mathcal{N}_t$ $(t = 1, ..., T)$,

- $q^{[n]} := \mathbb{P}[\Omega^{[n]}]$ is the probability to reach node n where $q^{[1]} = 1$,
- $(A_t^{[n]}, B_t^{[n]}, b_t^{[n]}, c_t^{[n]})$ is the realization of the random (A_t, B_t, b_t, c_t) at node n.

Because the constraints in `MSLP-`$\mathcal{P}$ do not need to hold for sets of measure zero, one can assume w.l.o.g. that $q^{[n]} > 0$, $n \in \mathcal{N}_t$ $(t = 1, ..., T)$. Now MSLP is rewritten as

$$\texttt{MSLP-}\mathcal{P}: \qquad \min_x \ \sum_{t=1}^T \sum_{n \in \mathcal{N}_t} q^{[n]} {c_t^{[n]}}^\top x_t^{[n]}$$

$$\text{subject to} \left\{ \begin{array}{rcll} A_1 x_1^{[1]} & = & b_1 & \\ B_t^{[n]} x_{t-1}^{[p_n]} + A_t^{[n]} x_t^{[n]} & = & b_t^{[n]} & , \ n \in \mathcal{N}_t \ \ (t = 2, ..., T) \\ x_t^{[n]} & \geq & 0 & , \ n \in \mathcal{N}_t \ \ (t = 1, ..., T) \ . \end{array} \right. \tag{3.1}$$

This problem can basically be solved exactly as a linear program. But the size of the LP exceeds rapidly the available storage capacity of computers, depending particularly on the number of modeled leaf nodes $n \in \mathcal{N}_T$. To be more precise, the dimension of the whole matrix involved is

$$\left(\sum_{t=1}^T m_t \cdot |\mathcal{N}_t|\right) \times \left(\sum_{t=1}^T n_t \cdot |\mathcal{N}_t|\right)$$

noting that A_t has dimension $(m_t \times n_t)$, $t = 1, ..., T$. At least, the matrix is sparse due to its staircase structure. Birge [4] developed a nested decomposition scheme for solving problems of this type. We also recommend the book of Kall/Mayer [16] to a reader interested in both a detailed description of the algorithm and a proof of its finiteness. The method is an extension of Benders' decomposition that is in fact a cutting plane method for special linear programs. An implementation is available due to Gassmann [11]. However, the subject of the present work is not to improve the capacity of an exact solution approach by assuming discrete distributions.

The dual of (3.1) - in terms of (2.10), Sect. 2.3 - will play an important part in our attempt of successive refinement. It reads as

$$\texttt{MSLP-}\mathcal{D}: \qquad \max_{u,s} \; \sum_{t=1}^{T} \sum_{n \in \mathcal{N}_t} q^{[n]} b_t^{[n]\top} u_t^{[n]}$$

$$\text{subject to} \left\{ \begin{array}{rcl} A_t^{[n]\top} u_t^{[n]} + \sum_{m \in C_n} \frac{q^{[m]}}{q^{[n]}} B_{t+1}^{[m]\top} u_{t+1}^{[m]} + s_t^{[n]} & = & c_t^{[n]}, \; n \in \mathcal{N}_t \\ & & (t = 1, ..., T-1) \\ A_T^{[n]\top} u_T^{[n]} + s_T^{[n]} & = & c_T^{[n]}, \; n \in \mathcal{N}_T \\ s_t^{[n]} & \geq & 0 \;\;, \; n \in \mathcal{N}_t \\ & & (t = 1, ..., T) \; . \end{array} \right. \tag{3.2}$$

Note that

$$\frac{q^{[m]}}{q^{[n]}} = \frac{\mathbb{P}[\Omega^{[m]}]}{\mathbb{P}[\Omega^{[n]}]} = \frac{\mathbb{P}[\Omega^{[m]} \cap \Omega^{[n]}]}{\mathbb{P}[\Omega^{[n]}]} = \mathbb{P}\big[\Omega^{[m]} \big| \, \Omega^{[n]}\big]$$

is the probability to reach node $m \in C_n$ given node n. The discrete counterpart of (2.12), Sect. 2.3, is as follows: if x is feasible in (3.1) and (u, s) is feasible in (3.2), then their objectives differ by

$$\sum_{t=1}^{T} \sum_{n \in \mathcal{N}_t} q^{[n]} c_t^{[n]\top} x_t^{[n]} - \sum_{t=1}^{T} \sum_{n \in \mathcal{N}_t} q^{[n]} b_t^{[n]\top} u_t^{[n]} = \sum_{t=1}^{T} \sum_{n \in \mathcal{N}_t} q^{[n]} \underbrace{s_t^{[n]\top} x_t^{[n]}}_{\geq 0} \geq 0, \tag{3.3}$$

and due to finite dimensionality, there is at least one feasible primal-dual pair $\big(\widehat{x}, (\widehat{u}, \widehat{s})\big)$ fulfilling

$$\widehat{s}_t^{[n]\top} \widehat{x}_t^{[n]} = 0, \;\; \forall n \in \mathcal{N}_t \; (t = 1, ..., T). \tag{3.4}$$

Remark 3.2
The primal-dual pair (3.1)-(3.2) has not precisely the identical form as (2.1)-(2.2) of Section 2.1. The usual dual in those terms is achieved by multiplying the equality constraints in (3.2) on both sides by $q^{[n]}$ followed by a substitution $\mathsf{u}_t^{[n]} := q^{[n]} u_t^{[n]}$ and $\mathsf{s}_t^{[n]} := q^{[n]} s_t^{[n]}$.

3.2 Approximating the distribution

Various theoretical (or heuristic) approaches for multistage stochastic programs - not only for the linear case - concern the approximation of the continuously distributed data by discrete ones, including empirical distributions resulting from Monte Carlo approaches. To be more precise, the theory is expedient for questions of stability under perturbation of the distribution. The statements can generally look as follows: suppose that $\xi = (\xi_2, ..., \xi_T) : \Omega \to \mathbb{R}^L$ is the vector of all random components occurring in the whole model where $\xi_t : \Omega \to \mathbb{R}^{l_t}$ is realized at time t and $l_2 + ... + l_T = L$. Then replace ξ with $\xi^{(k)} = (\xi_2^{(k)}, ..., \xi_T^{(k)}) : \Omega \to \mathbb{R}^L$ so that $\xi_2^{(k)}$ converges in distribution to ξ_2, denoted by $\xi_2^{(k)} \xrightarrow{\mathcal{D}} \xi_2$, and the conditional distributions satisfy

$$\xi^{(k)}_{t\,|\vec{\xi}^{(k)}_{t-1}} \xrightarrow{\mathcal{D}} \xi_{t\,|\,\vec{\xi}_{t-1}} \; (t = 3, ..., T)$$

as $k \to \infty$, where $\vec{\xi}_{t-1} := (\xi_2, ..., \xi_{t-1})$ and $\vec{\xi}^{(k)}_{t-1} := (\xi_2^{(k)}, ..., \xi_{t-1}^{(k)})$, $(t = 3, ..., T)$. One can show that this implies $\xi^{(k)} \xrightarrow{\mathcal{D}} \xi$, whereas the reverse does not need to be true (see e.g. Shapiro [29] for a discussion of this fact with regard to Monte-Carlo approaches leading to conditional sampling). Then, under appropriate assumptions (see e.g. Pennanen [24] for a rather general framework), the approximating optimal objective values $(\mathcal{Q}_1^{(k)})_{k \geq 0}$ converge to the unknown optimal value $\mathcal{Q}_1$ of the original problem. In addition, a statement can be included concerning the optimality of the accumulation points of the first-stage candidate solutions. Of course, a practical application is given whenever the approximating sequence $(\xi^{(k)})_{k \geq 0}$ is finite discrete, leading to a sequence of problem types (3.1). A related result has already been obtained in Olsen [23] and some similar approaches are given in Fiedler/Römisch [8], Pflug [25] and Pennanen [24], just to mention a few. In [25], the discretization error is expressed in terms of a Wasserstein distance d_W modulo a (non-computed or say non-computable) model constant, where

$$d_W(\xi^{(k)}, \xi) := \sup_{h \in \mathcal{H}} \left| E_{P^{(k)}} h(\xi^{(k)}) - E_P h(\xi) \right| \tag{3.5}$$

with P and $P^{(k)}$ being the probability measures in $\mathbb{R}^L$ induced by ξ and $\xi^{(k)}$, respectively, and $\mathcal{H}$ being the class of all bounded and Lipschitz continuous functions with Lipschitz constant 1, say, with respect to the maximum norm $|\cdot|_\infty$ in $\mathbb{R}^L$. The result below is certainly not surprising: it shows why crude discretization of the random data is probably not an appropriate way in terms of numerics in higher dimension $L = l_2 + ... + l_T$ because the growth of the number of leaf nodes in the approximating problems (3.1) can be enormous. For simplicity, the topic below is reduced to the uniform measure on the unit cube $[\![0,1]\!] \subset \mathbb{R}^L$.

Proposition 3.3
Suppose that $\xi : \Omega \to [\![0,1]\!] \subset \mathbb{R}^L$ is uniformly distributed on the unit cube and given $\xi^{(k)} : \Omega \to \mathbb{R}^L$ taking on $k \in \mathbb{N}$ different values. Then the Wasserstein distance (3.5) has a lower bound by

$$d_W(\xi^{(k)}, \xi) \geq \frac{1}{4}\left(\frac{1}{2k}\right)^{\frac{1}{L}}. \tag{3.6}$$

Hence, $d_W(\xi^{(k)}, \xi) \leq \varepsilon$ requires k to be larger than $\frac{1}{2}\left(\frac{1}{4\varepsilon}\right)^L$.

Proof. Suppose that $\xi^{(k)}$ takes on the values $\overline{z}^{(1)}, ..., \overline{z}^{(k)} \in \mathbb{R}^L$ with the probabilities $p_1, ..., p_k$. Let $S_i(\delta) := \{y \in \mathbb{R}^L \mid |y - \overline{z}^{(i)}|_\infty \leq \delta\}$ denote the ball around $\overline{z}^{(i)}$ with radius $\delta > 0$, $(i = 1, ..., k)$, and define $h_\delta : \mathbb{R}^L \longrightarrow \mathbb{R}$ by

$$h_\delta(y) := \begin{cases} |y - \overline{z}^{(\overline{i}(y))}|_\infty & \text{, if } y \in \bigcup_{i=1}^k S_i(\delta) \\ \delta & \text{, else} \end{cases}$$

where $\overline{i}(y) := \operatorname{argmin}_{i=1,...,k}\{i \mid |y - \overline{z}^{(i)}|_\infty = \min_{j=1,...,k} |y - \overline{z}^{(j)}|_\infty\}$. One easily verifies that h_δ is bounded and Lipschitz continuous with constant 1. This implies, for all $\delta > 0$,

$$d_W(\xi^{(k)}, \xi) \geq \left|E_{P^{(k)}} h_\delta(\xi^{(k)}) - \mathbb{E}_P h_\delta(\xi)\right| = \left|\textstyle\sum_{i=1}^k p_i h_\delta(\overline{z}^{(i)}) - \int_{[[0,1]]} h_\delta(z)dz\right|$$

where $h_\delta(\overline{z}^{(i)}) = 0$ holds, $(i = 1, ..., k)$. Let $\lambda^{(L)}$ denote the Lebesgue measure in $\mathbb{R}^L$, therefore $\lambda^{(L)}(S_i(\delta)) = (2\delta)^L$. It follows for all $\delta > 0$ that

$$\begin{aligned} d_W(\xi^{(k)}, \xi) &\geq \left|-\textstyle\int_{[[0,1]]} h_\delta(z)dz\right| = \textstyle\int_{[[0,1]]} h_\delta(z)dz \geq \delta\lambda^{(L)}\Big([[0,1]] \setminus \bigcup_{i=1}^k S_i(\delta)\Big) \\ &\geq \delta\Big(1 - \textstyle\sum_{i=1}^k \lambda^{(L)}\big(S_i(\delta)\big)\Big) = \delta\big(1 - k(2\delta)^L\big). \end{aligned}$$

In particular, inequality (3.6) is verified by choosing $\delta := \frac{1}{2}\left(\frac{1}{2k}\right)^{\frac{1}{L}}$. ☐

We conclude that, in general, approximating the original distribution in terms of the Wasserstein distance requires an exponential increase in atoms k with respect to the whole number $L = l_2 + ... + l_T$ of random components, independently of the way of discretization. On the other hand, a main handicap of crude discretization is the absence of an implementable and expedient measure of a discretization error in terms of the original optimization model. An approach using both-sided deterministic bounds is discussed in the following section.

3.3 Discrete bounding

Let us consider first the two-stage case to explain the concept behind discrete bounding. The deterministic formulation of the problem is therefore to determine

$$\begin{aligned} \mathcal{Q}_1 \quad := \quad & \inf_{\mathtt{y}_1 \in \mathbb{R}^{n_1}} c_1^\top \mathtt{y}_1 + \mathbb{E}\left[\mathcal{Q}_2(\mathtt{y}_1;\eta_2)\right] \\ & \text{s.t. } A_1 \mathtt{y}_1 = b_1 \\ & \qquad \mathtt{y}_1 \geq 0 \end{aligned} \tag{3.7}$$

where

$$\begin{aligned} \mathcal{Q}_2\big(\mathtt{y}_1;\eta_2\big) \quad := \quad & \min_{\mathtt{y}_2} c_2^\top \mathtt{y}_2 \\ & \text{s.t. } A_2 \mathtt{y}_2 = b_2 - B_2 \mathtt{y}_1 \\ & \qquad \mathtt{y}_2 \geq 0 \end{aligned} \tag{3.8}$$

and $\eta_2 := (A_2, B_2, b_2, c_2) : \Omega \to \mathbb{R}^{m_2 n_2 + m_2 n_1 + m_2 + n_2}$. Suppose that A_2 and c_2 are fixed, whereas $B_2 = \mathtt{B}_2(\xi_2)$ and $b_2 = \mathtt{b}_2(\xi_2)$, having affine linear mappings $\mathtt{B}_2 : \mathbb{R}^L \to \mathbb{R}^{m_2 \times n_1}$ and $\mathtt{b}_2 : \mathbb{R}^L \to \mathbb{R}^{m_2}$ in a random vector $\xi_2 : \Omega \to \mathbb{R}^L$. Hence one can write $\mathcal{Q}_2\big(\mathtt{y}_1;\xi_2\big)$ instead of $\mathcal{Q}_2\big(\mathtt{y}_1;\eta_2\big)$. The following result is a simple consequence of the basics in Section 2.1. For further details we refer to Kall/Wallace [14] and Birge/Louveaux [5].

Lemma 3.4
Let $\mathcal{B}_1 := \{\mathtt{y}_1 \in \mathbb{R}^{n_1} \mid A_1 \mathtt{y}_1 = b_1,\ \mathtt{y}_1 \geq 0\}$ be nonempty, and suppose that

A1) A_2 and c_2 are fixed,

A2) $\{\mathtt{v}_2 \mid A_2^\top \mathtt{v}_2 \leq c_2\} \neq \emptyset$,

A3) supp$\big\{\mathtt{b}_2(\xi_2) - \mathtt{B}_2(\xi_2)\mathtt{y}_1\big\} \subset A_2 \mathbb{R}_+^{n_2}$, $\forall \mathtt{y}_1 \in \mathcal{B}_1$.

Then the recourse function $\mathcal{Q}_2(\mathtt{y}_1;\bar{\xi}_2)$ is convex in $\bar{\xi}_2 \in$ conv$\big($supp$\{\xi_2\}\big)$ for all $\mathtt{y}_1 \in \mathcal{B}_1$, and Jensen's inequality yields

$$-\infty < \mathcal{Q}_2(\mathtt{y}_1;\mathbb{E}\xi_2) \leq \mathbb{E}\left[\mathcal{Q}_2(\mathtt{y}_1;\xi_2)\right] < +\infty\ ,\quad \forall \mathtt{y}_1 \in \mathcal{B}_1. \tag{3.9}$$

As a consequence one has

Proposition 3.5
Assume A1)-A3) of Lemma 3.4, and suppose that $\widehat{\mathrm{y}}_1$ *is a solution of*

$$\widehat{\mathcal{Q}}_1 := \min_{\mathrm{y}_1 \in \mathbb{R}^{n_1}} \left\{ c_1^\top \mathrm{y}_1 + \mathcal{Q}_2(\mathrm{y}_1; \mathbb{E}\xi_2) \;\middle|\; A_1 \mathrm{y}_1 = b_1,\ \mathrm{y}_1 \geq 0 \right\}.$$

In this case, bounds to the optimal objective value $\mathcal{Q}_1$ *apply by*

$$\widehat{\mathcal{Q}}_1 \leq \mathcal{Q}_1 \leq c_1^\top \widehat{\mathrm{y}}_1 + \mathbb{E}\left[\mathcal{Q}_2(\widehat{\mathrm{y}}_1; \xi_2)\right].$$

The computation of the lower bound $\widehat{\mathcal{Q}}_1$ is not that difficult in general because it can be computed as a linear program. However, the exact evaluation of $\mathbb{E}\left[\mathcal{Q}_2(\widehat{\mathrm{y}}_1; \xi_2)\right]$ is rarely possible; but suppose that the convex hull of the support of ξ_2 is given as the convex hull of finitely many points $\overline{z}^{(1)}, ..., \overline{z}^{(N)} \in \mathbb{R}^L$, i.e.,

$$\operatorname{conv}\big(\operatorname{supp}\{\xi_2\}\big) = \operatorname{conv}\{\overline{z}^{(1)}, ..., \overline{z}^{(N)}\}\ .$$

In this case, one proves that there are Borel-measurable functions $(\lambda_i : \mathbb{R}^L \to \mathbb{R})_{i=1,...,N}$ such that for all $z \in \operatorname{conv}\{\overline{z}^{(1)}, ..., \overline{z}^{(N)}\}$,

$$\lambda_i(z) \geq 0\ \forall i,\ \ \textstyle\sum_{i=1}^N \lambda_i(z) = 1 \quad \text{and } z = \sum_{i=1}^N \lambda_i(z) \overline{z}^{(i)}\ .$$

Furthermore, assuming A1)-A3) of Lemma 3.4, the recourse function $\mathcal{Q}_2(\widehat{\mathrm{y}}_1; \cdot)$ is convex on $\operatorname{conv}\{\overline{z}^{(1)}, ..., \overline{z}^{(N)}\}$. Thus, a computable upper bound is obtained by

$$\begin{aligned} \mathbb{E}\left[\mathcal{Q}_2(\widehat{\mathrm{y}}_1; \xi_2)\right] &= \mathbb{E}\left[\mathcal{Q}_2\big(\widehat{\mathrm{y}}_1; \textstyle\sum_{i=1}^N \lambda_i(\xi_2)\overline{z}^{(i)}\big)\ \right] \\ &\leq \mathbb{E}\left[\textstyle\sum_{i=1}^N \lambda_i(\xi_2)\mathcal{Q}_2\big(\widehat{\mathrm{y}}_1; \overline{z}^{(i)}\big)\right] \\ &= \textstyle\sum_{i=1}^N \big(\mathbb{E}\lambda_i(\xi_2)\big)\mathcal{Q}_2\big(\widehat{\mathrm{y}}_1; \overline{z}^{(i)}\big) \end{aligned} \tag{3.10}$$

provided that the complexity of evaluating the coefficients $(\mathbb{E}\lambda_i(\xi_2))_{i=1,...,N}$ is negligible and that the number N of extreme points is not to high. The best known example of this type of a bound is given by the *Edmundson-Madansky Upper Bound* (see e.g. Kall/Wallace [14] or Kall/Mayer [16] and references therein) where the support of ξ_2 is assumed to be contained in a rectangle having $N = 2^L$ extreme points. Then one has to subdivide the rectangle appropriately into smaller rectangles, changing over to conditional Jensen's bounds in order to improve the tightness. Efficient heuristics for refinement is discussed in Kall/Mayer [16] and an approved implementation, named `DAPPROX`, is due to the authors. If the number L of components of ξ_2 is low, say 6, sometimes up to 10, then the bounds get satisfyingly tight within a moderate effort as approved by many problem instances. The limit of this method is fast approaching since 2^L is exponential in L. This is probably the main reason why an extension of these upper bounding principles into a multistage framework seems to be somewhat infamous, also because in higher dimension, these upper bounds can be rather poor. Moreover, in the multi-stage case, the exact evaluation of the upper bound (3.10) is hardly possible since in this case, each value $\mathcal{Q}_2\big(\widehat{\mathrm{y}}_1; \overline{z}^{(i)}\big)$ can not be computed exactly. Analogously as bounding $\mathcal{Q}_1$ one can possibly replace first the third stage integrand with the numerically tractable Jensen's bound in order to obtain an upper bound for $\mathcal{Q}_2\big(\widehat{\mathrm{y}}_1; \overline{z}^{(i)}\big)$, and so on. A somewhat related approach is suggested in Frauendorfer [9] working instead with simplicial partitions and barycentric coordinates. The author also provides a proof of convergence under a "gridwidth to zero" assumption. The approach is addressed to a more general class of problems with randomness in both the right-hand side and the cost function.

3.4 Deterministic lower bounds by aggregation

In the previous section we have outlined that deterministic lower bounds of Jensen's type are easily obtained for two-stage problems with fixed (A_2, c_2). In the present section we analyze a kind of a multi-stage analogon for these lower bound types. Unfortunately, in the multi-stage case it is not sufficient to assume that $(A_t, c_t)_{t=2,\dots,T}$ or even $(A_t, B_t, c_t)_{t=2,\dots,T}$ is fixed in order to get lower bounds. From now on, similarly as in Wright [30], the extension is explained in terms of the infinite LP formulation of MSLP rather than of the deterministic formulation, because the recourse function values $\mathcal{Q}_t(\mathrm{y}_{t-1}, \vec{\eta}_t(\omega))$ cannot be exactly computed when $t < T$.

Recall that the stage information is given by a filtration $\mathcal{F} = (\mathcal{F}_t)_{t=1,\dots,T}$ in $(\Omega, \mathcal{A}, \mathbb{P})$. Now let $\widehat{\mathcal{F}} = (\widehat{\mathcal{F}}_t)_{t=1,..,T}$ be a subfiltration of $\mathcal{F}$, therefore $\widehat{\mathcal{F}}_t$ is a sub-σ-algebra of $\mathcal{F}_t$, $(t = 1, ..., T)$. This yields the diagram

$$\begin{array}{ccccccccc} & \widehat{\mathcal{F}}_1 & \subset & \widehat{\mathcal{F}}_2 & \subset & \cdots & \subset & \widehat{\mathcal{F}}_T & \\ & \parallel & & \cap & & & & \cap & \\ \{\emptyset, \Omega\} = & \mathcal{F}_1 & \subset & \mathcal{F}_2 & \subset & \cdots & \subset & \mathcal{F}_T & \subset \mathcal{A} \end{array} \quad . \tag{3.11}$$

Notations 3.6
If it is clear which subfiltration $\widehat{\mathcal{F}}$ is meant, then let $(\widehat{A}_1, \widehat{b}_1, \widehat{c}_1) := (A_1, b_1, c_1)$

and

$$(\widehat{A}_t, \widehat{B}_t, \widehat{b}_t, \widehat{c}_t) := \mathbb{E}\big[(A_t, B_t, b_t, c_t) \,\big|\, \widehat{\mathcal{F}}_t\big] \;\; (t = 2, ..., T).$$

Following from aggregation of all data and decisions in MSLP-$\mathcal{P}$ due to Wright [30], we consider

$$\texttt{MSLP-}\mathcal{P}(\widehat{\mathcal{F}}): \qquad \text{Minimize}_x \; \mathbb{E}\left[\textstyle\sum_{t=1}^T \widehat{c}_t^\top x_t\right]$$

$$\text{subject to} \begin{cases} \widehat{A}_1 x_1 & = & \widehat{b}_1 & \\ \widehat{B}_t x_{t-1} + \widehat{A}_t x_t & = & \widehat{b}_t \;\; \text{(a.s.)} & (t = 2, ..., T) \\ x_t & \geq & 0 \;\; \text{(a.s.)} & (t = 1, ..., T) \\ x_t & \in & L^2(\widehat{\mathcal{F}}_t; \mathbb{R}^{n_t}) & (t = 1, ..., T) \end{cases} . \tag{3.12}$$

The Model Assumptions 2.3 on page 12 ensure both the existence of conditional expectations and - in the case $\widehat{\mathcal{F}} = \mathcal{F}$ - the identity MSLP-$\mathcal{P}(\mathcal{F}) \,\widehat{=}\,$ MSLP-$\mathcal{P}$ (cf. (2.6)); therefore let us call MSLP-$\mathcal{P}(\mathcal{F})$ the *original problem*. If $\widehat{\mathcal{F}} := \mathcal{T} := \{\emptyset, \Omega\}_{t=1,...,T}$, then MSLP-$\mathcal{P}(\widehat{\mathcal{F}})$ is also called the *expected value problem* where all data are replaced by their expectation. A special attention has to be given to the dual of MSLP-$\mathcal{P}(\widehat{\mathcal{F}})$,

$$\texttt{MSLP-}\mathcal{D}(\widehat{\mathcal{F}}): \qquad \text{Maximize}_{u,s} \; \mathbb{E}\left[\textstyle\sum_{t=1}^T \widehat{b}_t^\top u_t\right]$$

$$\text{subject to} \begin{cases} \widehat{A}_t^\top u_t + \mathbb{E}[\widehat{B}_{t+1}^\top u_{t+1} \,|\, \widehat{\mathcal{F}}_t] + s_t & = & \widehat{c}_t \;\; \text{(a.s.)} & (t = 1, ..., T-1) \\ \widehat{A}_T^\top u_T + s_T & = & \widehat{c}_T \;\text{(a.s.)} & \\ s_t & \geq & 0 \;\; \text{(a.s.)} & (t = 1, ..., T) \\ u_t & \in & L^2(\widehat{\mathcal{F}}_t; \mathbb{R}^{m_t}) & (t = 1, ..., T) \\ s_t & \in & L^2(\widehat{\mathcal{F}}_t; \mathbb{R}^{n_t}) & (t = 1, ..., T) \end{cases} . \tag{3.13}$$

Consequently, MSLP-$\mathcal{D}(\mathcal{F})$ coincides with the original dual MSLP-$\mathcal{D}$ (cf. (2.10)). In the case that $\widehat{\mathcal{F}}$ is finitely generated, i.e., $\widehat{\mathcal{F}}_T$ is a finitely generated sub-σ-algebra of $\mathcal{A}$, the problems (3.12) and (3.13) can be represented in the nodal syntax of Section 3.1, resulting in the finite dimensional LP's (3.1) and (3.2).

The basic idea of aggregation/disaggregation is to construct an ascending chain $(\widehat{\mathcal{F}}^{(k)})_{k \geq 0}$ of finitely generated subfiltrations of $\mathcal{F}$, beginning with $\widehat{\mathcal{F}}^{(0)} = \{\emptyset, \Omega\}_{t=1,...,T}$, which leads to a sequence of approximating problems (3.12) and (3.13). Models are of special interest which permit the inheritance of feasibility from MSLP-$\mathcal{D}(\widehat{\mathcal{F}})$ to MSLP-$\mathcal{D}(\mathcal{F})$ in order to get lower bounds. This inheritance has also been investigated in Wright [30]. But contrary to those results, see [30] Lemma 8 and Thm. 9, we do not presume that the matrix structure $(B_t)_{t=2,...,T}$ is fixed. Hence in the case $T = 2$, our results are related to Lemma 3.4. First we need

Lemma 3.7
Let $\mathcal{G}$ and $\mathcal{H}$ be two independent sub-σ-algebras of $\mathcal{A}$, i.e. $\mathbb{P}[G \cap H] = \mathbb{P}[G]\mathbb{P}[H]$, $\forall G \in \mathcal{G}$, $H \in \mathcal{H}$. Furthermore, let $\widehat{\mathcal{G}} \subset \mathcal{G}$ be a sub-σ-algebra of $\mathcal{G}$. Then it follows that

a) $\mathbb{E}[X \,|\, \sigma(\widehat{\mathcal{G}}, \mathcal{H})] = \mathbb{E}[X \,|\, \widehat{\mathcal{G}}]$ *(a.s.),* $\forall X \in L^2(\mathcal{G}; \mathbb{R})$,

b) $\mathbb{E}[Y \,|\, \mathcal{G}] = \mathbb{E}[Y \,|\, \widehat{\mathcal{G}}]$ *(a.s.),* $\forall Y \in L^2\big(\sigma(\widehat{\mathcal{G}}, \mathcal{H}); \mathbb{R}\big)$.

Proof. By assumption, $X \in L^2(\mathcal{G}; \mathbb{R})$ is stochastically independent of $\mathcal{H}$. Thus, a) is a direct consequence of the behavior of conditional expectations with respect to independence (see e.g. Bauer [3]). We now prove a) $\Rightarrow$ b). Let $Y \in L^2\big(\sigma(\widehat{\mathcal{G}}, \mathcal{H}); \mathbb{R}\big)$ and

$$X := \big(\mathbb{E}[Y \,|\, \mathcal{G}] - \mathbb{E}[Y \,|\, \widehat{\mathcal{G}}]\big).$$

Because $X \in L^2(\mathcal{G}; \mathbb{R})$, it holds

$$\begin{aligned}
\mathbb{E}(X^2) &= \mathbb{E}\big(X\mathbb{E}[Y \,|\, \mathcal{G}]\big) - \mathbb{E}\big(X\mathbb{E}[Y \,|\, \widehat{\mathcal{G}}]\big) = \mathbb{E}\big(X\mathbb{E}[Y \,|\, \mathcal{G}]\big) \\
&= \mathbb{E}\big(\mathbb{E}[XY \,|\, \mathcal{G}]\big) = \mathbb{E}(XY) = \mathbb{E}\big(\mathbb{E}[XY \,|\, \sigma(\widehat{G}, \mathcal{H})]\big) \\
&= \mathbb{E}\big(\mathbb{E}[X \,|\, \sigma(\widehat{G}, \mathcal{H})]Y\big) \overset{\text{a)}}{=} \mathbb{E}\big(\underbrace{\mathbb{E}[X \,|\, \widehat{G}]}_{=0} Y\big) = 0,
\end{aligned}$$

that is $X = 0$ (a.s.). □

Lemma 3.8
Assume that $(A_t, c_t)_{t=2,...,T}$ is fixed, and suppose there are independent sub-σ-algebras $\mathcal{H}_1, ..., \mathcal{H}_T$ of $\mathcal{A}$ such that $\widehat{\mathcal{F}}$, $\mathcal{F}$ and $(B_t)_{t=2,...,T}$ satisfy

(i) $\mathcal{F}_t \subset \sigma(\mathcal{H}_1, ..., \mathcal{H}_t)$ $(t = 1, ..., T)$,

(ii) $B_t \sim \mathcal{H}_t$ $(t = 2, ..., T)$,

(iii) $\widehat{\mathcal{F}}_t \subset \sigma(\widehat{\mathcal{F}}_{t-1}, \mathcal{H}_t)$ $(t = 2, ..., T)$.

Then it holds

a) if $x = (x_1, ..., x_T)$ *is feasible in* MSLP-$\mathcal{P}(\mathcal{F})$*, then* $\widehat{x} := (x_1, \mathbb{E}[x_2|\widehat{\mathcal{F}}_2], ..., \mathbb{E}[x_T|\widehat{\mathcal{F}}_T])$ *is feasible in* MSLP-$\mathcal{P}(\widehat{\mathcal{F}})$*, and* x *and* $\widehat{x}$ *have an equal objective value in their problems;*

b) if (u, s) *is feasible in* MSLP-$\mathcal{D}(\widehat{\mathcal{F}})$*, then it is also feasible in* MSLP-$\mathcal{D}(\mathcal{F})$*, having an equal objective value.*

In accordance with Model Assumptions 2.3 we remark that $b_t \sim \mathcal{F}_t$ and $B_t \sim \mathcal{F}_t$ is assumed anyway. Thus, (ii) means $B_t \sim \mathcal{F}_t \cap \mathcal{H}_t$.

Proof.

a) Let $x = (x_1, ..., x_T)$ be feasible in MSLP-$\mathcal{P}(\mathcal{F})$ and

$$\widehat{x} = (\widehat{x}_1, ..., \widehat{x}_T) := \big(x_1, \mathbb{E}[x_2|\widehat{\mathcal{F}}_2], ..., \mathbb{E}[x_T|\widehat{\mathcal{F}}_T]\big).$$

We choose any $t \in \{2, ..., T\}$. Because of (i), $x_{t-1} \sim \mathcal{F}_{t-1}$ is stochastically independent of $\mathcal{H}_t$, therefore, Lemma 3.7 a) implies

$$\mathbb{E}\big[x_{t-1} \,|\, \sigma(\widehat{\mathcal{F}}_{t-1}, \mathcal{H}_t)\big] = \mathbb{E}[x_{t-1} \,|\, \widehat{\mathcal{F}}_{t-1}] = \widehat{x}_{t-1} \geq 0, \text{ (a.s.)}.$$

Thus,

$$\begin{aligned}
\mathbb{E}[B_t x_{t-1} \,|\, \widehat{\mathcal{F}}_t] &\overset{\text{(iii)}}{=} \mathbb{E}\Big[\mathbb{E}\big[B_t x_{t-1} \,|\, \sigma(\widehat{\mathcal{F}}_{t-1}, \mathcal{H}_t)\big] \,\Big|\, \widehat{\mathcal{F}}_t\Big] \\
&\overset{\text{(ii)}}{=} \mathbb{E}\Big[B_t \mathbb{E}\big[x_{t-1} \,|\, \sigma(\widehat{\mathcal{F}}_{t-1}, \mathcal{H}_t)\big] \,\Big|\, \widehat{\mathcal{F}}_t\Big] = \mathbb{E}\Big[B_t \widehat{x}_{t-1} \,\Big|\, \widehat{\mathcal{F}}_t\Big] \\
&= \mathbb{E}[B_t \,|\, \widehat{\mathcal{F}}_t]\widehat{x}_{t-1} = \widehat{B}_t \widehat{x}_{t-1}, \text{ (a.s.)}. \qquad (3.14)
\end{aligned}$$

The feasibility of x in MSLP-$\mathcal{P}(\mathcal{F})$ implies $B_t x_{t-1} + A_t x_t = b_t$ (a.s.). Since A_t is fixed, and together with (3.14), one concludes that

$$\begin{aligned}
\widehat{B}_t \widehat{x}_{t-1} + \widehat{A}_t \widehat{x}_t &= \mathbb{E}[B_t x_{t-1} \,|\, \widehat{\mathcal{F}}_t] + \mathbb{E}[A_t x_t \,|\, \widehat{\mathcal{F}}_t] \\
&= \mathbb{E}[B_t x_{t-1} + A_t x_t \,|\, \widehat{\mathcal{F}}_t] = \mathbb{E}[b_t \,|\, \widehat{\mathcal{F}}_t] = \widehat{b}_t, \text{ (a.s.)},
\end{aligned}$$

and therefore, $\widehat{x}$ is feasible in MSLP-$\mathcal{P}(\widehat{\mathcal{F}})$. In addition, because $(c_t)_{t=2,...,T}$ is also fixed, it holds

$$\mathbb{E}\left[\sum_{t=1}^{T} c_t^\top x_t\right] = \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}[c_t^\top x_t \,|\, \widehat{\mathcal{F}}_t]\right] = \mathbb{E}\left[\sum_{t=1}^{T} c_t^\top \mathbb{E}[x_t \,|\, \widehat{\mathcal{F}}_t]\right] = \mathbb{E}\left[\sum_{t=1}^{T} \widehat{c}_t^\top \widehat{x}_t\right],$$

and a) is proven.

b) Let $(u,s) = \big((u_1,s_1),...,(u_T,s_T)\big)$ be feasible in $\texttt{MSLP-}\mathcal{D}(\widehat{\mathcal{F}})$ and $t \in \{1,...,T-1\}$. Of course, u and s are also adapted to the filtration $\mathcal{F}$. Since A_t, c_t are fixed, it remains to show that

$$\mathbb{E}[\widehat{B}_{t+1}^\top u_{t+1} \,|\, \widehat{\mathcal{F}}_t] = \mathbb{E}[B_{t+1}^\top u_{t+1} \,|\, \mathcal{F}_t]. \tag{3.15}$$

The left-hand side is identical to

$$\begin{aligned}\mathbb{E}[\widehat{B}_{t+1}^\top u_{t+1} \,|\, \widehat{\mathcal{F}}_t] &= \mathbb{E}\Big[\mathbb{E}[B_{t+1}^\top \,|\, \widehat{\mathcal{F}}_{t+1}]u_{t+1} \,\Big|\, \widehat{\mathcal{F}}_t\Big] \\ &= \mathbb{E}\Big[\mathbb{E}[B_{t+1}^\top u_{t+1} \,|\, \widehat{\mathcal{F}}_{t+1}] \,\Big|\, \widehat{\mathcal{F}}_t\Big] = \mathbb{E}[B_{t+1}^\top u_{t+1} \,|\, \widehat{\mathcal{F}}_t].\end{aligned}$$

Because of (ii) and (iii), the product $B_{t+1}^\top u_{t+1}$ is surely measurable with respect to $\sigma(\widehat{\mathcal{F}}_t, \mathcal{H}_{t+1})$. And because of (i), we can apply Lemma 3.7 b), i.e., $\mathbb{E}[B_{t+1}^\top u_{t+1} \,|\, \widehat{\mathcal{F}}_t] = \mathbb{E}[B_{t+1}^\top u_{t+1} \,|\, \mathcal{F}_t]$. This completes the proof of (3.15). Comparing the objective values we have

$$\mathbb{E}\left[\sum_{t=1}^T \widehat{b}_t^\top u_t\right] = \mathbb{E}\left[\sum_{t=1}^T \mathbb{E}[b_t \,|\, \widehat{\mathcal{F}}_t]^\top u_t\right] = \mathbb{E}\left[\sum_{t=1}^T \mathbb{E}[b_t^\top u_t \,|\, \widehat{\mathcal{F}}_t]\right] = \mathbb{E}\left[\sum_{t=1}^T b_t^\top u_t\right].$$

□

Corollary 3.9
Under the assumptions of Lemma 3.8 it follows that

$$\begin{array}{ccc} \inf\big(\textit{MSLP-}\mathcal{P}(\widehat{\mathcal{F}})\big) & \leq & \inf\big(\textit{MSLP-}\mathcal{P}(\mathcal{F})\big) \\ \rotatebox{90}{$\leq$} & & \rotatebox{90}{$\leq$} \\ \sup\big(\textit{MSLP-}\mathcal{D}(\widehat{\mathcal{F}})\big) & \leq & \sup\big(\textit{MSLP-}\mathcal{D}(\mathcal{F})\big) \end{array} .$$

Proof. Lemma 3.8 a) yields the upper, whereas b) the lower inequality. The weak duality (cf. (2.13)) completes the diagram. □

The assumptions of Lemma 3.8 always hold in the following two examples. We remark that (i)-(iii) rely on an interaction of $\mathcal{F}$, $\widehat{\mathcal{F}}$ and $(B_t)_{t=2,...,T}$.

Example 3.10 (Two-stage case, only B_2, b_2 being random)
Let $T = 2$ and (A_2, c_2) be fixed. Set $\mathcal{H}_1 := \{\emptyset, \Omega\}$ and $\mathcal{H}_2 := \mathcal{F}_2$, which are trivially independent. Thus, (i)-(iii) are satisfied.

Example 3.11 (Trivial subfiltration, only $(b_t)_{t=2,...,T}$ being random)
Let $(A_t, B_t, c_t)_{t=2,...,T}$ be fixed and $\widehat{\mathcal{F}} := \{\emptyset, \Omega\}_{t=1,...,T}$. Then (i)-(iii) hold by choosing $\mathcal{H}_1 := \mathcal{F}_T$ and $\mathcal{H}_t := \{\emptyset, \Omega\}(t = 2, ..., T)$.

Remarks 3.12

1) *An example is given in Wright [30] where (iii) in Lemma 3.8 does not apply (the author chooses a subfiltration* $\widehat{\mathcal{F}}$ *with* $\widehat{\mathcal{F}}_1 = \widehat{\mathcal{F}}_2 = \{\emptyset, \Omega\}$ *and* $\widehat{\mathcal{F}}_3 \subset \mathcal{F}_2$*). Another example is computed in Fusek et al. [10] with the absence of independent* $\mathcal{H}_t$*'s. As a result of both examples, no lower bound is obtained by aggregation. Both examples contain a three-stage problem where only* b_2, b_3 *are random. Thus, it is misleadingly mentioned in Birge/Louveaux [5], p. 356, that there is no need of an independent structure for lower bounds as long as only* $(b_t)_{t=2,...,T}$ *varies. In general, this merely holds true if* $\widehat{\mathcal{F}} := \{\emptyset, \Omega\}_{t=1,...,T}$ *(see Example 3.11).*

2) *In Wright [30] Lemma 8 and Thm. 9, the technical assumption*

$$\mathbb{E}[X \mid \widehat{\mathcal{F}}_t] = \mathbb{E}[X \mid \widehat{\mathcal{F}}_T] \text{ (a.s.)}, \ \forall X \in L^2(\mathcal{F}_t; \mathbb{R}) \ \ (t = 2, ..., T) \qquad (3.16)$$

replaces (i) and (iii) of Lemma 3.8 in order to get the same statements. But in addition, $(B_t)_{t=2,...,T}$ *must be assumed to be fixed (or at least* $B_t \sim \widehat{\mathcal{F}}_t$*, but that is inexpedient anyway).*

For the sake of completeness we mention

Corollary 3.13

The assumptions (i),(iii) of Lemma 3.8 imply the property (3.16).

Proof. Let $t \in \{2, ..., T\}$, and given $X \in L^2(\mathcal{F}_t; \mathbb{R})$. From (iii) it follows that $\widehat{\mathcal{F}}_T \subset \sigma(\widehat{\mathcal{F}}_t, \mathcal{H}_{t+1}, ..., \mathcal{H}_T)$, and by (i), $X \sim \mathcal{F}_t$ is independent of $\sigma(\mathcal{H}_{t+1}, ..., \mathcal{H}_T)$. Thus, Lemma 3.7 a) implies

$$\begin{aligned} \mathbb{E}[X \mid \widehat{\mathcal{F}}_T] &= \mathbb{E}\big[\mathbb{E}[X \mid \sigma(\widehat{\mathcal{F}}_t, \mathcal{H}_{t+1}, ..., \mathcal{H}_T)] \,\big|\, \widehat{\mathcal{F}}_T\big] \\ &= \mathbb{E}\big[\mathbb{E}[X \mid \widehat{\mathcal{F}}_t] \,\big|\, \widehat{\mathcal{F}}_T\big] = \mathbb{E}[X \mid \widehat{\mathcal{F}}_t]. \end{aligned}$$

□

In order to have a practical application of Lemma 3.8, a special setting is at hand when the accretion of information at each stage t is stochastically independent of the past.

Theorem 3.14

Suppose that $\xi_2, ..., \xi_T$ *are stochastically independent random vectors such that*

V1) $\mathcal{F}_t = \sigma(\xi_2, ..., \xi_t)$ *(*$t = 2, ..., T$*),*

V2) $(A_t, c_t)_{t=2,...,T}$ *is fixed,* *V3)* $B_t \sim \sigma(\xi_t)$, $b_t \sim \sigma(\xi_2, ..., \xi_t)$, $(t = 2, ..., T)$,

V4) $\mathtt{MSLP}\text{-}\mathcal{P}(\mathcal{F})$ *is feasible,*

V5) $\mathtt{MSLP}\text{-}\mathcal{D}(\mathcal{T})$ *is feasible where* $\mathcal{T} := \{\emptyset, \Omega\}_{t=1,...,T}$.

Given two finitely generated subfiltrations $\widehat{\mathcal{F}}$ *and* $\widehat{\mathcal{F}}'$ *with* $\widehat{\mathcal{F}} \subset \widehat{\mathcal{F}}' \subset \mathcal{F}$ *and*

$$\widehat{\mathcal{F}}_t \subset \sigma(\widehat{\mathcal{F}}_{t-1}, \xi_t)\ ,\ \ \mathcal{F}'_t \subset \sigma(\widehat{\mathcal{F}}'_{t-1}, \xi_t) \quad (t = 2, ..., T). \tag{3.17}$$

Then it follows that every feasible solution of $\mathtt{MSLP}\text{-}\mathcal{D}(\widehat{\mathcal{F}})$ *is feasible in both* $\mathtt{MSLP}\text{-}\mathcal{D}(\widehat{\mathcal{F}}')$ *and* $\mathtt{MSLP}\text{-}\mathcal{D}(\mathcal{F})$*, in particular,*

$$\begin{array}{ccccc} \min\big(\mathtt{MSLP}\text{-}\mathcal{P}(\widehat{\mathcal{F}})\big) & \leq & \min\big(\mathtt{MSLP}\text{-}\mathcal{P}(\widehat{\mathcal{F}}')\big) & \leq & \inf\big(\mathtt{MSLP}\text{-}\mathcal{P}(\mathcal{F})\big) \\ \| & & \| & & \rotatebox{90}{$\leq$} \\ \max\big(\mathtt{MSLP}\text{-}\mathcal{D}(\widehat{\mathcal{F}})\big) & \leq & \max\big(\mathtt{MSLP}\text{-}\mathcal{D}(\widehat{\mathcal{F}}')\big) & \leq & \sup\big(\mathtt{MSLP}\text{-}\mathcal{D}(\mathcal{F})\big)\ . \end{array} \tag{3.18}$$

Proof. By choosing $\mathcal{H}_1 := \{\emptyset, \Omega\}$ and $\mathcal{H}_t := \sigma(\xi_t)$ $(t = 2, ..., T)$, all conditions of Lemma 3.8 are satisfied for both $\widehat{\mathcal{F}}$ and $\widehat{\mathcal{F}}'$. Each feasible (u, s) in $\mathtt{MSLP}\text{-}\mathcal{D}(\widehat{\mathcal{F}})$ is certainly also adapted to $\widehat{\mathcal{F}}'$. Let $\widehat{B}'_t := \mathbb{E}[B_t \,|\, \widehat{\mathcal{F}}'_t]$. By (3.15) it has been shown that $\mathbb{E}[B_{t+1}^\top u_{t+1} \,|\, \mathcal{F}_t]$ is identical to $\mathbb{E}[\widehat{B}_{t+1}^\top u_{t+1} \,|\, \widehat{\mathcal{F}}_t]$ (a.s.), therefore, it is also identical to $\mathbb{E}[\widehat{B}'^\top_{t+1} u_{t+1} \,|\, \widehat{\mathcal{F}}'_t]$. Hence we have $\mathbb{E}[\widehat{B}_{t+1}^\top u_{t+1} \,|\, \widehat{\mathcal{F}}_t] = \mathbb{E}[\widehat{B}'^\top_{t+1} u_{t+1} \,|\, \widehat{\mathcal{F}}'_t]$. This means that (u, s) is likewise feasible in $\mathtt{MSLP}\text{-}\mathcal{D}(\widehat{\mathcal{F}}')$, in fact with equal objective value. Hence $\sup\big(\mathtt{MSLP}\text{-}\mathcal{D}(\widehat{\mathcal{F}})\big) \leq \sup\big(\mathtt{MSLP}\text{-}\mathcal{D}(\widehat{\mathcal{F}}')\big)$ holds.
In particular, V5) implies the feasibility of all dual problems in (3.18), and in connection with V4), Lemma 3.8 a) ensures the same property for all the primal problems. Finally, due to the finiteness of $\widehat{\mathcal{F}}$ and $\widehat{\mathcal{F}}'$, strong LP duality holds for the aggregated problems (see also (3.4)), and the diagram (3.18) is completely verified. □

The measurability structure V1)-V3) of the above theorem is illustrated in Figure 3.2. Our way of composing and refining subfiltrations is not addressed until Chapter 5.

Randomness in the cost vectors

One might also analyze models which preserve primal feasibility from $\mathtt{MSLP}\text{-}\mathcal{P}(\widehat{\mathcal{F}})$ to the original $\mathtt{MSLP}\text{-}\mathcal{P}(\mathcal{F})$ resulting in upper bounds. For

c_1	c_2	$c_3 \cdots$	c_{T-1}	c_T	
A_1					b_1
$B_2(\xi_2)$	A_2				$b_2(\xi_2)$
	$B_3(\xi_3)$	A_3			$b_3(\xi_2, \xi_3)$
		$\ddots$	$\ddots$		$\vdots$
			$B_T(\xi_T)$	A_T	$b_T(\xi_2, ..., \xi_T)$

Figure 3.2: *Permitted stochastics in data to get lower bounds by aggregation. In addition, $\xi_2, ..., \xi_T$ must be stochastically independent even if $(B_t)_{t=2,...,T}$ is totally fixed.*

this purpose, the necessary assumptions are slightly asymmetric to those of Lemma 3.8, namely, the whole structure $(A_t, B_t, b_t)_{t=2,...,T}$ must be fixed, but there is no need for stochastically independent increments in the process $(c_t)_{t=2,...,T}$. We also refer to Wright [30] for this case. There is at least one special situation where randomness in both the right-hand side $(b_t)_{t=2,...,T}$ and the cost vectors $(c_t)_{t=2,...,T}$ can be combined in such a way that the setting V1)-V3) of Theorem 3.14 is valid after a transformation. Suppose that $\mathcal{F}_t = \sigma(\xi_2, ..., \xi_t)$ $(t = 2, ..., T)$, having stochastically independent vectors

$$\xi_2 := \begin{pmatrix} \eta_2 \\ \xi'_2 \end{pmatrix}, \cdots, \xi_T := \begin{pmatrix} \eta_T \\ \xi'_T \end{pmatrix}$$

where $\eta_t : \Omega \to (0, +\infty)$ is a positive random variable, possibly correlated with the random vector ξ'_t, and having that both η_t and the reciprocal $\frac{1}{\eta_t}$ are essentially bounded. For $t = 2, ..., T$, assume that (A_t, B_t) is fixed, $b_t \sim \sigma(\xi'_2, ..., \xi'_t)$ and the cost vector is given as $c_t := (\prod_{r=1}^{t} \eta_r)\mathtt{c}_t$ where $\eta_1 := 1$ and $\mathtt{c}_t \in \mathbb{R}^{n_t}$. A substitution $x'_t := (\prod_{r=1}^{t} \eta_r)x_t$ results in the equivalence

$$\begin{aligned} & B_t x_{t-1} + A_t x_t = b_t \ , \ x_t \geq 0, \ \text{(a.s.)}, \ x_t \in L^2(\mathcal{F}_t; \mathbb{R}^{n_t}) \\ \iff \ & \eta_t B_t x'_{t-1} + A_t x'_t = \textstyle\prod_{r=1}^{t} \eta_r b_t \ , \ x'_t \geq 0, \ \text{(a.s.)}, \ x'_t \in L^2(\mathcal{F}_t; \mathbb{R}^{n_t}) \ . \end{aligned}$$

Hence the original problem $\mathtt{MSLP}\text{-}\mathcal{P}(\mathcal{F})$ is rewritten here equivalently as

$$\mathtt{MSLP}\text{-}\mathcal{P}'(\mathcal{F}) : \qquad \text{Minimize}_{x'} \ \mathbb{E}\left[\textstyle\sum_{t=1}^{T} \mathtt{c}_t^\top x'_t\right]$$

$$\text{subject to} \begin{cases} A_1 x_1' &= b_1 & \\ B_t' x_{t-1}' + A_t x_t' &= b_t' \quad \text{(a.s.)} & (t = 2, ..., T) \\ x_t' &\geq 0 \quad \text{(a.s.)} & (t = 1, ..., T) \\ x_t' &\in L^2(\mathcal{F}_t; \mathbb{R}^{n_t}) & (t = 1, ..., T) \end{cases} \tag{3.19}$$

where for $t = 2, ..., T$,

- $B_t' := \eta_t B_t \sim \sigma(\eta_t) \subset \sigma(\xi_t)$,
- $b_t' := \prod_{r=2}^{t} \eta_r b_t \sim \sigma\big((\eta_2, \xi_2'), ..., (\eta_t, \xi_t')\big) = \sigma(\xi_2, ..., \xi_t)$.

Now the setting V1)-V3) of Theorem 3.14 is applicable for MSLP-$\mathcal{P}'(\mathcal{F})$.

3.5 Statistical upper bounds by recursions

In the previous section, mainly in Theorem 3.14, lower bounding principles have been analyzed and extended from Wright [30]. However, this kind of aggregation does not provide upper bounds for the unknown objective value. Now we shall explain our main idea to obtain a feasible primal policy in the original model leading to (statistical) upper bounds. Suppose that $\widehat{\mathcal{F}}$ is a finitely generated subfiltration of $\mathcal{F}$, take for instance $\widehat{\mathcal{F}} := \{\emptyset, \Omega\}_{t=1,...,T}$, and given an optimal solution $\widehat{x} = (\widehat{x}_1, ..., \widehat{x}_T)$ and $(\widehat{u}, \widehat{s}) = \big((\widehat{u}_1, \widehat{s}_1), ..., (\widehat{u}_T, \widehat{s}_T)\big)$ of MSLP-$\mathcal{P}(\widehat{\mathcal{F}})$ and MSLP-$\mathcal{D}(\widehat{\mathcal{F}})$, respectively. Hence the complementarity $\mathbb{E}[\widehat{s}^\top \widehat{x}] = 0$ is satisfied, or equivalently (cf. also (3.4)),

$$\widehat{s}_t^\top \widehat{x}_t = 0 \text{ (a.s.)} \quad (t = 1, ..., T).$$

Under appropriate assumptions, e.g. as those of Theorem 3.14, $(\widehat{u}, \widehat{s})$ is at least feasible in the original dual MSLP-$\mathcal{D}(\mathcal{F})$. However, $\widehat{x} = (\widehat{x}_1, ..., \widehat{x}_T)$ is rarely feasible in MSLP-$\mathcal{P}(\mathcal{F})$. Hence we ask for a continuation $(\overline{x}_2, ..., \overline{x}_T)$ of the first stage candidate solution $\widehat{x}_1$ so that $\overline{x} = (\widehat{x}_1, \overline{x}_2, ..., \overline{x}_T)$ might turn out to be feasible and near-optimal in the original problem MSLP-$\mathcal{P}(\mathcal{F})$. Because $\widehat{s}_1^\top \widehat{x}_1 = 0$ holds true, the duality gap of such a pair $\big(\overline{x}, (\widehat{u}, \widehat{s})\big)$ is expressed as (see (2.12))

$$\begin{aligned} \mathbb{E}[\mathbf{c}^\top \overline{x}] - \mathbb{E}[\mathbf{b}^\top \widehat{u}] &= \mathbb{E}\left[\textstyle\sum_{t=1}^{T} c_t^\top \overline{x}_t\right] - \mathbb{E}\left[\textstyle\sum_{t=1}^{T} b_t^\top \widehat{u}_t\right] \\ &= \mathbb{E}\left[\textstyle\sum_{t=1}^{T} \widehat{s}_t^\top \overline{x}_t\right] = \mathbb{E}\left[\textstyle\sum_{t=2}^{T} \widehat{s}_t^\top \overline{x}_t\right] = \mathbb{E}[\widehat{s}^\top \overline{x}]. \end{aligned} \tag{3.20}$$

When the decision $\overline{x}_{t-1}(\omega)$ is established for each $\omega \in \Omega$, beginning with $\overline{x}_1(\omega) = \widehat{x}_1\ \forall \omega \in \Omega$, we define the pointwise LP solution set

$$\Psi_t(\omega) := \underset{\mathrm{y}_t \in \mathbb{R}^{n_t}}{\operatorname{argmin}} \left\{ \widehat{s}_t(\omega)^\top \mathrm{y}_t \,\middle|\, A_t(\omega)\mathrm{y}_t = b_t(\omega) - B_t(\omega)\overline{x}_{t-1}(\omega),\ \mathrm{y}_t \geq 0 \right\} \subset \mathbb{R}^{n_t}. \tag{3.21}$$

Because this set is frequently (or say mostly) not a singleton, we pick up

$$\overline{x}_t(\omega) := \operatorname{argmin}_{\mathrm{y}_t \in \Psi_t(\omega)} |\mathrm{y}_t - \widehat{x}_t(\omega)| \in \mathbb{R}^{n_t}. \tag{3.22}$$

We will see in Section 4.2 that - under appropriate assumptions - the resulting policy $\overline{x} = (\overline{x}_1, \overline{x}_2, ..., \overline{x}_T)$ is well-defined and feasible in $\texttt{MSLP-}\mathcal{P}(\mathcal{F})$. After having evaluated the lower bound

$$\mathrm{LB} := \max\big(\texttt{MSLP-}\mathcal{D}(\widehat{\mathcal{F}})\big) = \mathbb{E}[\mathbf{b}^\top \widehat{u}] \leq \sup\big(\texttt{MSLP-}\mathcal{D}(\mathcal{F})\big) \leq \inf\big(\texttt{MSLP-}\mathcal{P}(\mathcal{F})\big), \tag{3.23}$$

an upper bound to the optimal objective value is given by

$$\mathrm{UB} := \mathrm{LB} + \mathbb{E}[\widehat{s}^\top \overline{x}] = \mathbb{E}[\mathbf{b}^\top \widehat{u}] + \mathbb{E}[\widehat{s}^\top \overline{x}] \overset{(3.20)}{=} \mathbb{E}[\mathbf{c}^\top \overline{x}] \geq \inf\big(\texttt{MSLP-}\mathcal{P}(\mathcal{F})\big) \tag{3.24}$$

where the last inequality follows from the feasibility of $\overline{x}$ in $\texttt{MSLP-}\mathcal{P}(\mathcal{F})$. An unbiased estimator of the upper bound is obtained due to an i.i.d. sample of size N by

$$\overline{\mathrm{UB}} := \mathrm{LB} + \tfrac{1}{N}\textstyle\sum_{j=1}^{N} \widehat{s}(\omega_j)^\top \overline{x}(\omega_j) = \mathrm{LB} + \tfrac{1}{N}\textstyle\sum_{j=1}^{N} \left[\textstyle\sum_{t=2}^{T} \widehat{s}_t(\omega_j)^\top \overline{x}_t(\omega_j)\right].$$

The statistical upper bound $\overline{\mathrm{UB}}$ is obviously almost surely larger than the deterministic lower bound LB because they differ by the nonnegative sum $\frac{1}{N}\sum_{j=1}^{N} \left[\sum_{t=1}^{T} \widehat{s}_t(\omega_j)^\top \overline{x}_t(\omega_j)\right]$. Of course, if $\widehat{s}_t^\top \overline{x}_t = 0$ (a.s.), $t = 2, ..., T$, or equivalently, if almost all optimal LP objective values in (3.21) are zero, then the primal-dual pair $\big(\overline{x}, (\widehat{u}, \widehat{s})\big)$ is optimal. In the next chapter we analyze a worst-case behavior of the complementarity variable $\widehat{s}^\top \overline{x} : \Omega \to [0, +\infty)$ depending on aggregation errors. This permits thereafter to propose an implementable approximation scheme on the basis of local refinements; as a basic rule in doing so, disaggregation of $\widehat{\mathcal{F}}$ will take place *at the most* in regions where (simulated) values $\widehat{s}(\omega)^\top \overline{x}(\omega) > \varepsilon > 0$ are observed above a prescribed tolerance $\varepsilon > 0$.

4 Validation of the recursive policy

Section 3.4 has analyzed sufficient conditions to obtain feasible dual solutions $(\widehat{u},\widehat{s}) = \big((\widehat{u}_1,\widehat{s}_1),...,(\widehat{u}_T,\widehat{s}_T)\big)$ in the original dual problem MSLP-$\mathcal{D}(\mathcal{F})$ by replacing the original filtration $\mathcal{F}$ by a coarser one $\widehat{\mathcal{F}} \subset \mathcal{F}$. Based on aggregated solutions the basic idea to define recursively a primal policy $\overline{x} = (\overline{x}_1,...,\overline{x}_T)$ adapted to $\mathcal{F}$ has been proposed in the previous section. The purpose of the present chapter is to give suitable conditions which ensure that $\overline{x}$ is well-defined and near-optimal in MSLP-$\mathcal{P}(\mathcal{F})$. The first section deals with a unique selection of non-unique LP solutions in the context of sensitivity analysis. After that in Section 4.2, the results will be applied to the subproblems (3.22) in combination with (3.21). The aim is to investigate a worst-case behavior of the complementarity variable $\widehat{s}^\top\overline{x} : \Omega \to [0,+\infty)$ depending on the aggregation error and adding up to the duality gap $\mathbb{E}[\widehat{s}^\top\overline{x}]$. We also make slightly weaker assumptions in order to specify solely a worst-case behavior of $\mathbb{E}[\widehat{s}^\top\overline{x}]$.

4.1 Parametrized projections onto LP solution sets

Let us consider the parametrized LP

$$\gamma = \gamma(b,c) = \min_{x\in\mathbb{R}^n}\big\{c^\top x \;\big|\; Ax = b,\ x \geq 0\big\} \tag{4.1}$$

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$. Furthermore, let $\Psi = \Psi(b,c) \subset \mathbb{R}^n$ denote the solution set of (4.1). The strong LP duality, cf. Prop. 2.1 b), implies that

$$\Psi \neq \emptyset \Longleftrightarrow b \in A\mathbb{R}^n_+,\ c \in (A^\top\mathbb{R}^m + \mathbb{R}^n_+)\ .$$

Whenever $\Psi \neq \emptyset$ is not a singleton (in this case one says that the dual of (4.1) is degenerate) we pick up

$$\overline{x} := \operatorname{argmin}_{x\in\Psi}|x-y|$$

where $y \in \mathbb{R}^n$ is a further parameter. Because $|x-y| = \big((x-y)^\top(x-y)\big)^{\frac{1}{2}}$, the selection $\overline{x}$ results from the Euclidean projection of y onto the LP solution set; therefore, $\overline{x}$ is well-defined and unique since Ψ is closed and convex. We remark that in the setting of (3.21) and (3.22), (A,b,c,y) corresponds to $\big(A_t(\omega), b_t(\omega) - B_t(\omega)\overline{x}_{t-1}(\omega), \widehat{s}_t(\omega), \widehat{x}_t(\omega)\big)$, whereas x and $\overline{x}$ correspond to $\mathtt{y}_t$ and $\overline{x}_t(\omega)$, respectively.

The goal of this section is to prove that $\overline{x}$ is Lipschitz continuous with respect to perturbations of b and y and, moreover, that the Lipschitz constant $L = L(A)$ is independent of the objective vector c. A fundamental result concerning a unique selection of non-unique LP solutions has been derived in *Mangasarian/Meyer* [19]:

Proposition 4.1 (cf. [19], Cor. 2)
Let $\Psi \neq \emptyset$ be the solution set of (4.1) and $f : \mathbb{R}^n \to \mathbb{R}$ be strictly convex. If $\min_{x\in\Psi} f(x)$ is solvable with the (unique) solution $\overline{x}$, then $\overline{x}$ is also the unique solution of

$$\min_x \left\{ c^\top x + \sigma f(x) \;\middle|\; Ax = b,\ x \geq 0 \right\}$$

for all sufficiently small $\sigma > 0$.

Because $|x - y|^2$ is strictly convex in x, it immediately follows

Corollary 4.2
Let $\Psi \neq \emptyset$ be the solution set of (4.1), $\overline{x} := \mathrm{argmin}_{x\in\Psi}|x - y|$ and

$$\overline{x}_\sigma := \mathrm{argmin}_x \left\{ c^\top x + \frac{\sigma}{2}|x - y|^2 \;\middle|\; Ax = b,\ x \geq 0 \right\} \ ,\ \sigma > 0\ . \tag{4.2}$$

Then $\overline{x} = \overline{x}_\sigma$ holds for all sufficiently small $\sigma > 0$.

Thus, for analyzing a certain Lipschitz behavior of $\overline{x}$ we can first do the same for $\overline{x}_\sigma$. This leads us shortly into the theory of convex quadratic programs under linear constraints with regard to the problem

$$\min_{z\in\mathbb{R}^n} \{ e^\top z + z^\top Q z \mid Dz \geq d \} \tag{4.3}$$

where $Q \in \mathbb{R}^{n\times n}$ is positive semidefinite and $D \in \mathbb{R}^{m\times n}$. *Klatte/Thiere* [17] have proved

Proposition 4.3 (cf. [17], Thm. 4.2)
Let $\Phi(d) \subset \mathbb{R}^n$ denote the optimal solution set of (4.3) depending on the right-hand side $d \in \mathbb{R}^m$. Then $\Phi : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is Lipschitzian on its domain $\mathrm{dom}\Phi := \{d \mid \Phi(d) \neq \emptyset\}$, i.e., there exists an $L = L(D, Q, e) > 0$ such that

$$\mathit{dist}\big(\Phi(d), \Phi(d')\big) \leq L|d - d'|\ ,\ \forall d, d' \in \mathrm{dom}\Phi, \tag{4.4}$$

where $\mathit{dist}(X, Y) := \max\left\{ \max_{x\in X} \min_{y\in Y} |x - y|, \max_{y\in Y} \min_{x\in X} |y - x| \right\}$ denotes the Hausdorff distance between $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^n$.

Of course, if Q is even positive definite, then $\Phi(d)$ is a singleton for all $d \in \mathrm{dom}\Phi$ and (4.4) reads as $|\Phi(d) - \Phi(d')| \le L|d - d'|$. This is a main argument in the proof of

Lemma 4.4
Given $A \in \mathbb{R}^{m\times n}$ and $\rho(\cdot,\cdot,\cdot) : A\mathbb{R}^n_+ \times (A^\top\mathbb{R}^m + \mathbb{R}^n_+) \times \mathbb{R}^n \longrightarrow \mathbb{R}^n$,

$$\rho(b, c, y) := \operatorname{argmin}_{x\in\Psi(b,c)} |x - y| \tag{4.5}$$

where $\Psi(b, c)$ is the optimal LP solution set of (4.1). The mapping $\rho(\cdot,\cdot,\cdot)$ is Borel measurable on its domain. Moreover, there is an $L = L(A)$ such that

$$|\rho(b', c, y') - \rho(b, c, y)| \le L\left(|b' - b| + |y' - y|\right) ,$$

$\forall b, b' \in A\mathbb{R}^n_+$, $\forall c \in (A^\top\mathbb{R}^m + \mathbb{R}^n_+)$, $\forall y, y' \in \mathbb{R}^n$.

We emphasize that $\rho(b,\cdot,y)$ is not continuous let alone Lipschitz in c. However, the constant L does not depend on c.

Proof. For any $b \in A\mathbb{R}^n_+$, $c, y \in \mathbb{R}^n$ and $\sigma > 0$, let

$$\rho_\sigma(b, c, y) := \operatorname{argmin}_x \left\{ c^\top x + \frac{\sigma}{2}|x - y|^2 \;\middle|\; Ax = b,\ x \ge 0 \right\}.$$

A straightforward computation shows that

$$c^\top x + \frac{\sigma}{2}|x - y|^2 = \frac{\sigma}{2}\left|x - \left(y - \frac{1}{\sigma}c\right)\right|^2 + c^\top\left(y - \frac{1}{2\sigma}c\right)$$

where the last term is independent of x. Hence $\rho_\sigma(b, c, y)$ is the projection of $\left(y - \frac{1}{\sigma}c\right)$ onto the feasible set, therefore, it is well-defined and unique. The substitution $z := x - \left(y - \frac{1}{\sigma}c\right)$ results in

$$\rho_\sigma(b, c, y) = \left(y - \frac{1}{\sigma}c\right) + \operatorname*{argmin}_z \left\{ z^\top z \;\middle|\; Az = b - A\left(y - \frac{1}{\sigma}c\right),\ z \ge -\left(y - \frac{1}{\sigma}c\right) \right\}.$$

We may express the equality constraints '$Az = h$' as '$Az \ge h$' in combination with '$-Az \ge -h$'. Now we can apply Proposition 4.3 with $e := 0$, $Q := I$ and

$$D := \begin{pmatrix} A \\ -A \\ I \end{pmatrix}, \quad d := \begin{pmatrix} b - A\left(y - \frac{1}{\sigma}c\right) \\ -b + A\left(y - \frac{1}{\sigma}c\right) \\ -\left(y - \frac{1}{\sigma}c\right) \end{pmatrix}$$

where I is the $(n \times n)$-identity matrix. It is simple to show that the Lipschitz property (4.4) now holds for $\rho_\sigma(b, c, y)$. In particular, there is an $L = L(D, Q, e) = L(A)$ with

$$|\rho_\sigma(b', c', y') - \rho_\sigma(b, c, y)| \leq L \left(|b' - b| + \frac{1}{\sigma}|c' - c| + |y' - y| \right), \tag{4.6}$$

$\forall b, b' \in A\mathbb{R}^n_+$, $\forall c, c', y, y' \in \mathbb{R}^n$. Because $\rho_\sigma(\cdot, \cdot, \cdot)$ is Lipschitzian, it is also continuous and therefore Borel measurable.
Corollary 4.2 now has two implications. First one concludes that $\rho(\cdot, \cdot, \cdot) = \lim_{\sigma \downarrow 0} \rho_\sigma(\cdot, \cdot, \cdot)$ on the domain of $\rho(\cdot, \cdot, \cdot)$, i.e. pointwise convergence holds. Hence the limit function $\rho(\cdot, \cdot, \cdot)$ is also Borel measurable. On the other hand, if $\sigma > 0$ is sufficiently small (depending on b, c, y and b', y'), Corollary 4.2 yields in combination with (4.6),

$$|\rho(b', c, y') - \rho(b, c, y)| = |\rho_\sigma(b', c, y') - \rho_\sigma(b, c, y)| \leq L \left(|b' - b| + |y' - y| \right).$$

Note that the left and right term is independent of $\sigma > 0$, whereas the right term is even independent of $c \in (A^\top \mathbb{R}^m + \mathbb{R}^n_+)$. This completes the proof. □

As a by-product we obtain

Corollary 4.5
Given a sub-σ algebra $\mathcal{G}$ of $\mathcal{A}$. For any matrix $A \in \mathbb{R}^{m \times n}$ let

$$\tilde{b} \in L^2(\mathcal{G}, A\mathbb{R}^n_+), \ \tilde{c} : \Omega \to (A^\top \mathbb{R}^m + \mathbb{R}^n_+), \ \tilde{y} \in L^2(\mathcal{G}, \mathbb{R}^n),$$

where $\tilde{c}$ is measurable with respect to $\mathcal{G}$. Then it holds that $\rho(\tilde{b}, \tilde{c}, \tilde{y}) \in L^2(\mathcal{G}, \mathbb{R}^n)$.

Proof. Because $\rho(\cdot, \cdot, \cdot)$ is Borel measurable due to Lemma 4.4, the interconnection $\rho(\tilde{b}, \tilde{c}, \tilde{y}) : \Omega \to \mathbb{R}^n$ is measurable with respect to $\mathcal{G}$. Furthermore, since $\tilde{c} \in (A^\top \mathbb{R}^m + \mathbb{R}^n_+)$ (a.s.), it obviously holds $\rho(0, \tilde{c}, 0) = 0$ (a.s.). Therefore, the constant $L = L(A)$ in Lemma 4.4 yields

$$|\rho(\tilde{b}, \tilde{c}, \tilde{y})| = |\rho(0, \tilde{c}, 0) - \rho(\tilde{b}, \tilde{c}, \tilde{y})| \leq L\big(|\tilde{b}| + |\tilde{y}|\big) \ \text{(a.s.)}.$$

Together with $\tilde{b}$ and $\tilde{y}$, the left side is square integrable because the right side is independent of $\tilde{c}$. □

For $b \in A\mathbb{R}^n_+$, $c \in (A^\top \mathbb{R}^m + \mathbb{R}^n_+)$ and $y \in \mathbb{R}^n$, the evaluation of $\rho(b, c, y)$ requires first to compute the optimal LP objective value

$$\gamma := \min_x \{ c^\top x \mid Ax = b, \ x \geq 0 \}$$

and to solve thereafter the QP

$$\rho(b,c,y) = \text{argmin}_x \big\{ (x-y)^T(x-y) \;\big|\; Ax = b,\ c^\top x = \gamma,\ x \geq 0 \big\}. \tag{4.7}$$

Note that (4.7) is near-infeasible because every small perturbation of γ into $\gamma' < \gamma$ leads to an infeasible set. But these kinds of problems will appear as very small dimensional subproblems and one can solve them numerically stable e.g. with `GAMS/MINOS` [6]. Otherwise, we would recommend to replace (4.7) with a regularized LP (4.2). Sometimes, one can even derive an explicit representation of $\rho(\cdot,\cdot,\cdot)$ as in the following setting.

The simple recourse case

Let us suppose that $A := (I, -I) \in \mathbb{R}^{m \times 2m}$ where I is the $(m \times m)$-identity. Then the projection (4.7) is separable into simplest problems of dimension 2 because also the Euclidean norm square

$$\begin{aligned}(x-y)^\top(x-y) &= (x_1-y_1)^\top(x_1-y_1) + (x_2-y_2)^\top(x_2-y_2) \\ &= \sum_{i=1}^{m} \left[(x_{1i}-y_{1i})^2 + (x_{2i}-y_{2i})^2 \right]\end{aligned}$$

is additive separable. Thus, without loss of generality, we may assume $m = 1$. Then it holds $A\mathbb{R}^{2m}_+ = \mathbb{R}$ and $A^\top \mathbb{R}^m + \mathbb{R}^{2m}_+ = \big\{ \left(\begin{smallmatrix} c_1 \\ c_2 \end{smallmatrix} \right) \in \mathbb{R}^2 \big| \; c_1 + c_2 \geq 0 \big\}$. For any $b \in \mathbb{R}$ and $c = (c_1, c_2)^\top \in \mathbb{R}^2$, $c_1 + c_2 \geq 0$, the optimal LP objective value is given by

$$\gamma = \gamma(b,c) = \min_x \{ c_1 x_1 + c_2 x_2 \mid x_1 - x_2 = b,\ x_1, x_2 \geq 0 \} = \begin{cases} c_1 b, & \text{if } b \geq 0 \\ c_2 b, & \text{if } b \leq 0 \end{cases}$$

where the optimal LP solution set is written as

$$\Psi(b,c) = \begin{cases} \begin{cases} \left\{ \begin{pmatrix} b \\ 0 \end{pmatrix} \right\}, \text{ if } b \geq 0 \\ \left\{ \begin{pmatrix} 0 \\ -b \end{pmatrix} \right\}, \text{ if } b \leq 0 \end{cases} & , \text{ if } c_1 + c_2 > 0 \\ \left\{ \begin{pmatrix} x_1 \\ x_1 - b \end{pmatrix} \;\middle|\; x_1 \geq \max\{0, b\} \right\} & , \text{ if } c_1 + c_2 = 0 \end{cases} .$$

This set is therefore a singleton whenever $c_1 + c_2 > 0$, and in this case, the projection of $y \in \mathbb{R}^2$ onto $\Psi(b,c)$ is trivial. Otherwise, if $c_1 + c_2 = 0$, a straightforward computation results in

$$\rho(b,c,y) = \begin{cases} \begin{pmatrix} b \\ 0 \end{pmatrix} & \text{, if } \; b \geq \max\{0, y_1+y_2\} \\ \begin{pmatrix} 0 \\ -b \end{pmatrix} & \text{, if } \; b \leq \min\{0, -(y_1+y_2)\} \\ \frac{1}{2}\begin{pmatrix} y_1+y_2+b \\ y_1+y_2-b \end{pmatrix} & \text{, otherwise} \end{cases} .$$

Only one case results in a non-vertex of the feasible set, namely, if $c_1+c_2=0$ and $-(y_1+y_2) < b < (y_1+y_2)$. However, ignoring this case would be disastrous in our concept because (near-) optimal decision policies in MSLP generally are not allowed to be selected as sequences of extreme points (see also the comments in the introductory example of Section 1.1).

4.2 Worst-case behavior of the complementarity variable

In this section, the quality of the recursively defined policy proposed in Section 3.5 is analyzed. For this purpose we have to include an assumption on MSLP called "relatively complete fixed recourse" (RCR), where the expression 'fixed' refers to fixed matrices $A_2, ..., A_T$. There is probably no standardized definition of RCR in the literature concerning multistage decision problems because RCR can be expressed in algebraic terms or in a purely verbal sense. The latter reads as follows: *"MSLP is RCR if there are feasible first-stage decisions and, furthermore, if every decision policy, feasible by the time $t < T$, has a feasible continuation at stage $t+1$"*. We shall keep with this formulation, also, because it is consistent with the usual RCR setting of the two-stage case (see also Lemma 3.4). Our formal definition is

Definition 4.6 (RCR)
Let $\mathcal{B}_1 := \{x_1 \in \mathbb{R}^{n_1} \mid A_1 x_1 = b_1,\ x_1 \geq 0\}$ and, for $t = 2, ..., T$,

$$\mathcal{B}_t := \Big\{ (x_1, ..., x_t) \;\Big|\; \begin{array}{rcl} (x_1, ..., x_{t-1}) & \in & \mathcal{B}_{t-1}, \\ B_t x_{t-1} + A_t x_t & = & b_t \text{ (a.s.)} \\ x_t & \geq & 0 \text{ (a.s.)} \\ x_t & \in & L^2(\mathcal{F}_t, \mathbb{R}^{n_t}) \end{array} \Big\}.$$

`MSLP-`$\mathcal{P}(\mathcal{F})$ is said to have relatively complete fixed recourse (RCR) if $(A_t)_{t=2,...,T}$ is fixed and

(i) $b_1 \in A_1\mathbb{R}_+^{n_1}$,

(ii) *supp* $\{b_t - B_t x_{t-1}\} \subset A_t\mathbb{R}_+^{n_t}$, $\forall(x_1, ..., x_{t-1}) \in \mathcal{B}_{t-1} \quad (t = 2, ..., T)$.

Note that $\mathcal{B}_T$ coincide with the feasible set of `MSLP-`$\mathcal{P}(\mathcal{F})$. The problem is obviously RCR whenever $A_t\mathbb{R}_+^{n_t} = \mathbb{R}^{m_t}$ $(t = 1, ..., T)$.

Theorem 4.7
For a given finitely generated subfiltration $\widehat{\mathcal{F}}$ *of* $\mathcal{F}$*, assume that*

A1) `MSLP-`$\mathcal{P}(\mathcal{F})$ *is RCR;*

A2) $\widehat{x} = (\widehat{x}_1, ..., \widehat{x}_T)$ *solves* `MSLP-`$\mathcal{P}(\widehat{\mathcal{F}})$ *(cf. (3.12));*

A3) $(\widehat{u}, \widehat{s}) = \big((\widehat{u}_1, \widehat{s}_1), ..., (\widehat{u}_T, \widehat{s}_T)\big)$ *solves* `MSLP-`$\mathcal{D}(\widehat{\mathcal{F}})$ *(cf. (3.13)).*

Define recursively

- $\overline{x}_1 := \widehat{x}_1$,
- *for* $t = 2, ..., T$, $\forall\omega \in \Omega$,

$$\Psi_t(\omega) := \underset{\mathrm{y}_t\in\mathbb{R}^{n_t}}{\operatorname{argmin}} \left\{\widehat{s}_t(\omega)^\top \mathrm{y}_t \,\middle|\, A_t\mathrm{y}_t = b_t(\omega) - B_t(\omega)\overline{x}_{t-1}(\omega),\ \mathrm{y}_t \geq 0\right\} \subset \mathbb{R}^{n_t},$$
$$\overline{x}_t(\omega) := \underset{\mathrm{y}_t\in\Psi_t(\omega)}{\operatorname{argmin}}\ |\mathrm{y}_t - \widehat{x}_t(\omega)| \in \mathbb{R}^{n_t}.$$

Then the policy $\overline{x} := (\overline{x}_1, ..., \overline{x}_T)$ *is feasible in* `MSLP-`$\mathcal{P}(\mathcal{F})$ *and there is a model constant* $L \in \mathbb{R}_+$*, independent of* $\widehat{\mathcal{F}}$*, such that*

$$|\overline{x} - \widehat{x}| \leq L\textstyle\sum_{t=2}^{T} \Big(|b_t - \widehat{b}_t| + |B_t - \widehat{B}_t||\widehat{x}_{t-1}|\Big) \text{ (a.s.).} \tag{4.8}$$

We remark that L is solely deterministic in (4.8), whereas all other quantities are given by Euclidean vector- or matrix norms, respectively, producing a random outcome.

Proof. According to Definition 4.6 one shows by an induction argument on $t = 1, ..., T$ that $(\overline{x}_1, ..., \overline{x}_t) \in \mathcal{B}_t$. One has obviously $\overline{x}_1 = \widehat{x}_1 \in \{\mathrm{y}_1 \in \mathbb{R}^{n_1} \mid \widehat{A}_1\mathrm{y}_1 = \widehat{b}_1,\ \mathrm{y}_1 \geq 0\} = \mathcal{B}_1$. Assume that the hypothesis is true for $t-1$, i.e. $(\overline{x}_1, ..., \overline{x}_{t-1}) \in \mathcal{B}_{t-1}$. RCR implies that $\operatorname{supp}\{b_t - B_t\overline{x}_{t-1}\} \subset A_t\mathbb{R}_+^{n_t}$; therefore, one can apply Corollary 4.5 using $\mathcal{G} := \mathcal{F}_t$, $\widetilde{b} := (b_t - B_t\overline{x}_{t-1})$, $\tilde{y} := \widehat{x}_t$ and $\tilde{c} := \widehat{s}_t$. Note that $\widetilde{b} \in L^2(\mathcal{F}_t; \mathbb{R}^{m_t})$ holds by the induction hypothesis, whereas $\widetilde{y} \in L^2(\mathcal{F}_t; \mathbb{R}^{m_t})$ is trivially given since $\widetilde{y} \sim \widehat{\mathcal{F}}_t$ takes on a finite number of values (a.s.). Moreover, one has $\mathrm{supp}\{\widehat{s}_t\} \subset \mathbb{R}^{n_t}_+ \subset (A_t^\top \mathbb{R}^{m_t} + \mathbb{R}^{n_t}_+)$ (a.s.). Corollary 4.5 states therefore that $\overline{x}_t \in L^2(\mathcal{F}_t; \mathbb{R}^{n_t})$, and hence, $\overline{x} = (\overline{x}_1, ..., \overline{x}_t) \in \mathcal{B}_t$. This completes the proof of the induction. In particular, $\overline{x} = (\overline{x}_1, ..., \overline{x}_T) \in \mathcal{B}_T$ means that $\overline{x}$ is feasible in $\mathtt{MSLP}\text{-}\mathcal{P}(\mathcal{F})$.

It remains to prove (4.8). In accordance with Lemma 4.4 we use the notation of the projections on the LP solution sets induced by the individual recourse matrices $A_2, ..., A_T$,

$$\rho_t(\cdot, \cdot, \cdot) : A_t\mathbb{R}^{n_t}_+ \times (A_t^\top \mathbb{R}^{m_t} + \mathbb{R}^{n_t}_+) \times \mathbb{R}^{n_t} \longrightarrow \mathbb{R}^{n_t} \quad (t = 2, ..., T)$$

where for $(\mathtt{d}_t, \mathtt{s}_t, \mathtt{z}_t) \in A_t\mathbb{R}^{n_t}_+ \times (A_t^\top \mathbb{R}^{m_t} + \mathbb{R}^{n_t}_+) \times \mathbb{R}^{n_t}$, $t = 2, ..., T$,

$$\rho_t(\mathtt{d}_t, \mathtt{s}_t, \mathtt{z}_t) := \mathrm{argmin}_{\mathtt{x}_t \in \Psi_t(\mathtt{d}_t, \mathtt{s}_t)} |\mathtt{x}_t - \mathtt{z}_t|$$

and $\Psi_t(\mathtt{d}_t, \mathtt{s}_t) := \mathrm{argmin}_{\mathtt{x}_t \in \mathbb{R}^{n_t}} \{\mathtt{s}_t^\top \mathtt{x}_t \mid A_t \mathtt{x}_t = \mathtt{d}_t,\ \mathtt{x}_t \geq 0\}$. Because the pair $\big(\widehat{x}, (\widehat{u}, \widehat{s})\big)$ is assumed to be optimal in the aggregated (and finite discrete) problems, the complementarity $\mathbb{E}[\widehat{s}^\top \widehat{x}] = 0$ holds. Since also $\widehat{x} \geq 0$ and $\widehat{s} \geq 0$ (a.s.), this means equivalently that $\widehat{s}_t^\top \widehat{x}_t = 0$ (a.s.), $(t = 1, ..., T)$. Thus one has obviously the identities

$$\widehat{x}_t = \rho_t\big(\widehat{b}_t - \widehat{B}_t \widehat{x}_{t-1}, \widehat{s}_t, \widehat{x}_t\big) \text{ (a.s.)}, \quad (t = 2, ..., T).$$

Furthermore, the recursively defined policy $\overline{x}$ of the theorem is given by $\overline{x}_1 = \widehat{x}_1$ and

$$\overline{x}_t = \rho_t\big(b_t - B_t \overline{x}_{t-1}, \widehat{s}_t, \widehat{x}_t\big) \text{ (a.s.)}, \quad (t = 2, ..., T).$$

For each $t \in \{2, ..., T\}$ there is an $L_t = L(A_t)$ due to Lemma 4.4 such that

$$\begin{aligned}
|\overline{x}_t - \widehat{x}_t| &= \Big|\rho_t\big(\, b_t - B_t \overline{x}_{t-1}, \widehat{s}_t, \widehat{x}_t\big) - \rho_t\big(\widehat{b}_t - \widehat{B}_t \widehat{x}_{t-1}, \widehat{s}_t, \widehat{x}_t\big)\Big| \\
&\leq L_t \Big|(b_t - B_t \overline{x}_{t-1}) - (\widehat{b}_t - \widehat{B}_t \widehat{x}_{t-1})\Big| \\
&= L_t \Big|(b_t - \widehat{b}_t) + (\widehat{B}_t - B_t)\widehat{x}_{t-1} + B_t(\widehat{x}_{t-1} - \overline{x}_{t-1})\Big| \\
&\leq L_t \Big(|b_t - \widehat{b}_t| + |B_t - \widehat{B}_t||\widehat{x}_{t-1}| + |B_t||\overline{x}_{t-1} - \widehat{x}_{t-1}|\Big) \\
&\leq \overline{L}_t \Big(|b_t - \widehat{b}_t| + |B_t - \widehat{B}_t||\widehat{x}_{t-1}| + |\overline{x}_{t-1} - \widehat{x}_{t-1}|\Big), \text{(a.s.)}, \qquad (4.9)
\end{aligned}$$

where we let $\overline{L}_t := \max\big\{1, L_t \max\{1, \mathrm{ess\ sup}|B_t|\}\big\}$ (≥ 1). Now, by inserting inequality (4.9) backwards for all $2 \leq r < t$ and taking into account that $\overline{x}_1 = \widehat{x}_1$, the result is

$$|\overline{x}_t - \widehat{x}_t| \leq \big(\textstyle\prod_{r=2}^t \overline{L}_r\big) \sum_{r=2}^t \Big(|b_r - \widehat{b}_r| + |B_r - \widehat{B}_r||\widehat{x}_{r-1}|\Big) \text{ (a.s.)}. \qquad (4.10)$$

The right-hand side of (4.10) is monotonically increasing in t. Hence it follows that

$$\begin{aligned} |\overline{x} - \widehat{x}| &\leq \textstyle\sum_{t=2}^{T} |\overline{x}_t - \widehat{x}_t| \\ &\leq \underbrace{(T-1)\left(\textstyle\prod_{r=2}^{T} \overline{L}_t\right)}_{=:L} \textstyle\sum_{r=2}^{T} \left(|b_r - \widehat{b}_r| + |B_r - \widehat{B}_r||\widehat{x}_{r-1}|\right), \text{ (a.s.)} . \end{aligned}$$

This completes the proof of Theorem 4.7. □

Lemma 4.8
Assume that $b_2, ..., b_T$ are essentially bounded, $(A_t)_{t=2,...,T}$ is fixed and

$$\{\mathtt{z}_t \in \mathbb{R}^{n_t} \mid A_t \mathtt{z}_t = 0,\ \mathtt{z}_t \geq 0\} = \{0\}\ ,\ (t = 1, ..., T). \tag{4.11}$$

Then there is a model constant $\beta \in \mathbb{R}_+$ such that $|x| \leq \beta$ (a.s.) for all feasible x in MSLP-$\mathcal{P}(\widehat{\mathcal{F}})$ and for all subfiltrations $\widehat{\mathcal{F}}$ of $\mathcal{F}$ (including the case $\widehat{\mathcal{F}} = \mathcal{F}$).

Proof. Let $x = (x_1, ..., x_T)$ be feasible in MSLP-$\mathcal{P}(\widehat{\mathcal{F}})$. Assuming (4.11), by Proposition 2.1 d) there are constants $\alpha_1 := L(A_1), ..., \alpha_t := L(A_T) \in \mathbb{R}_+$ such that

$$|x_1| \leq \alpha_1 |b_1| =: \beta_1$$

and

$$|x_t| \leq \alpha_t \left|\widehat{b}_t - \widehat{B}_t x_{t-1}\right| \text{ (a.s.) } (t = 2, ..., T).$$

Recall that $B_2, ..., B_T$ are essentially bounded by model assumptions on page 12. Now it holds inductively for $t = 2, ..., T$,

$$\begin{aligned} |x_t| &\leq \alpha_t \left|\widehat{b}_t - \widehat{B}_t x_{t-1}\right| \leq \alpha_t \big(|\widehat{b}_t| + |\widehat{B}_t||x_{t-1}|\big) \\ &\leq \alpha_t \big(\text{ess sup}|\widehat{b}_t| + \text{ess sup}|\widehat{B}_t||x_{t-1}|\big) \\ &\leq \alpha_t \big(\text{ess sup}|b_t| + \text{ess sup}|B_t||x_{t-1}|\big) \\ &\leq \alpha_t \big(\text{ess sup}|b_t| + \text{ess sup}|B_t|\beta_{t-1}\big) =: \beta_t,\ \text{(a.s.)}. \end{aligned}$$

Hence $|x| \leq \sum_{t=1}^{T} |x_t| \leq \sum_{t=1}^{T} \beta_t := \beta$, (a.s.). □

Remark 4.9
As already mentioned earlier, the original problem MSLP-$\mathcal{P}(\mathcal{F})$ is surely RCR whenever $A_t \mathbb{R}_+^{n_t} = \mathbb{R}^{m_t}$ $(t = 1, ..., T)$. But we remark that this latter property would be incompatible with (4.11). That is to say, for every matrix $A \in \mathbb{R}^{m \times n}$ holds the implication

$$\{\mathtt{z} \in \mathbb{R}^n \mid A\mathtt{z} = 0,\ \mathtt{z} \geq 0\} = \{0\} \quad \Longrightarrow \quad A\mathbb{R}_+^n \subsetneqq \mathbb{R}^m\ , \tag{4.12}$$

because by contraposition there is an $\mathtt{x} \geq 0$ with $A\mathtt{x} = -Ae$ where $e := (1, ..., 1)^\top$; hence $\mathtt{z} := \mathtt{x} + e > 0$ and $A\mathtt{z} = 0$.

Next we establish a slight modification of Definition 4.6.

Definition 4.10 (RCR^o)
We say that MSLP-$\mathcal{P}(\mathcal{F})$ *is RCR^o if* $(A_t)_{t=2,\dots,T}$ *is fixed and*

(i) $b_1 \in int(A_1\mathbb{R}_+^{n_1})$,

(ii) $supp\, \{b_t - B_t x_{t-1}\} \subset int(A_t\mathbb{R}_+^{n_t})\ ,\ \forall (x_1,...,x_{t-1}) \in \mathcal{B}_{t-1} \quad (t = 2,...,T)$.

Of course, RCR^o implies RCR, but the difference is rather marginal. We ask whether RCR^o is sufficient to ensure that the optimal solutions of MSLP-$\mathcal{D}(\widehat{\mathcal{F}})$ are essentially bounded, uniformly with respect to the choice $\widehat{\mathcal{F}} \subset \mathcal{F}$. The answer is positive for the two-stage case (a similar subject is investigated in Birge/Louveaux [5], Exe. 11, p. 103). More generally we state

Lemma 4.11
Assume that $c_2,...,c_T$ *are essentially bounded and*
B1) MSLP-$\mathcal{P}(\mathcal{F})$ *is RCR^o;*

B2) $\mathcal{F}_t = \sigma(\xi_2,...,\xi_t)$ $(t = 2,...,T)$ *where* $\xi_2,....,\xi_T$ *are independent random vectors;*

B3) $supp\{(B_2,b_2),...,(B_T,b_T)\} \subset \mathbb{R}^{\sum_{t=2}^T m_t(n_{t-1}+1)}$ *is convex.*
Then there is a model constant $\gamma \in \mathbb{R}_+$ *such that* $\left|\begin{pmatrix}\widehat{u}\\ \widehat{s}\end{pmatrix}\right| \le \gamma$ *(a.s.)*

- $\forall (\widehat{u},\widehat{s}) \in argmax\big($MSLP-$\mathcal{D}(\widehat{\mathcal{F}})\big)$,
- $\forall$ *finitely generated subfiltrations* $\widehat{\mathcal{F}}$ *of* $\mathcal{F}$ *satisfying* $\widehat{\mathcal{F}}_t \subset \sigma(\widehat{\mathcal{F}}_{t-1},\xi_t)$ $(t = 2,...,T)$.

Proof. Suppose that $\widehat{\mathcal{F}}$ is a finitely generated subfiltration with the assumed property, and given $(\widehat{u},\widehat{s}) = \big((\widehat{u}_1,\widehat{s}_1),...,(\widehat{u}_T,\widehat{s}_T)\big) \in \text{argmax}\big($MSLP-$\mathcal{D}(\widehat{\mathcal{F}})\big)$ together with any $\widehat{x} = (\widehat{x}_1,...,\widehat{x}_T) \in \text{argmin}\big($MSLP-$\mathcal{P}(\widehat{\mathcal{F}})\big)$. First we show that B1)-B3) ensure that

$$\text{supp}\{\widehat{b}_t - \widehat{B}_t\widehat{x}_{t-1}\} \subset \text{int}\big(A_t\mathbb{R}_+^{n_t}\big)\ (t = 2,...,T). \tag{4.13}$$

(Note: if $T \ge 3$, then this property follows not immediately from B1)! See also later Example 4.14). Because RCR^o obviously implies RCR, the setting A1)-A3) of Theorem 4.7 is applicable; in particular, based on $\widehat{x}$ and $\widehat{s}$, the recursively defined policy $\overline{x} = (\widehat{x}_1,\overline{x}_2,...,\overline{x}_T)$ in Theorem 4.7 is feasible in MSLP-$\mathcal{P}(\mathcal{F})$, and by (4.10) we proved that

$$|\widehat{x}_t - \overline{x}_t| \le \left(\textstyle\prod_{r=2}^t \overline{L}_r\right) \textstyle\sum_{r=2}^t \left(|\widehat{b}_r - b_r| + |\widehat{B}_r - B_r||\widehat{x}_{r-1}|\right) \text{ (a.s.)},\ (t = 2,...,T)\ .$$

For a given $t \in \{2, ..., T\}$ it follows that

$$\begin{aligned}\left|(\widehat{b}_t - \widehat{B}_t\widehat{x}_{t-1}) - (b_t - B_t\overline{x}_{t-1})\right| &= |A_t\widehat{x}_t - A_t\overline{x}_t| \le |A_t||\widehat{x}_t - \overline{x}_t| \\ &\le \lambda \textstyle\sum_{r=2}^{t} \left(|\widehat{b}_r - b_r| + |\widehat{B}_r - B_r|\right), \text{ (a.s.)} \qquad (4.14)\end{aligned}$$

where $\lambda := |A_t| \left(\prod_{r=2}^{t} \overline{L}_r\right) \max\left\{1, |\widehat{x}_1|, \text{ess sup}|\widehat{x}_2|, ..., \text{ess sup}|\widehat{x}_{t-1}|\right\} \in \mathbb{R}_+$. Let $\mathbf{B}_t := \left((B_2, b_2), ..., (B_t, b_t)\right)$ and $\widehat{\mathbf{B}}_t := \left((\widehat{B}_2, \widehat{b}_2), ..., (\widehat{B}_t, \widehat{b}_t)\right)$ be represented as random vectors. The sum in (4.14) describes a norm $|\cdot|_*$ in the image space $\mathbb{R}^{\sum_{r=2}^{t} m_r(n_{r-1}+1)}$ of $\mathbf{B}_t$, this means that

$$\left|(\widehat{b}_t - \widehat{B}_t\widehat{x}_{t-1}) - (b_t - B_t\overline{x}_{t-1})\right| \le \lambda|\widehat{\mathbf{B}}_t - \mathbf{B}_t|_* \text{ (a.s.)}. \qquad (4.15)$$

Due to the assumption B2) and because $\widehat{\mathcal{F}}_r \subset \sigma(\widehat{\mathcal{F}}_{r-1}, \xi_r)$ $(r = 2, ..., t)$, we can apply Corollary 3.4. Hence the property $\widehat{\mathbf{B}}_t = \mathbb{E}[\mathbf{B}_t \mid \widehat{\mathcal{F}}_t]$ (a.s.) holds. Therefore, assumption B3) implies that $\widehat{\mathbf{B}}_t \in \text{supp}\{\mathbf{B}_t\}$ (a.s.). Thus by (4.15) and B1), it follows that

$$\text{supp}\left\{\widehat{b}_t - \widehat{B}_t\widehat{x}_{t-1}\right\} \subset \text{supp}\left\{b_t - B_t\overline{x}_{t-1}\right\} \subset \text{int}\left(A_t\mathbb{R}_+^{n_t}\right).$$

Since $t \in \{2, ..., T\}$ was arbitrary, the assertion (4.13) is verified.

Next we define

$$\widehat{d}_1 := b_1 \ , \ \ \widehat{d}_t := \widehat{b}_t - \widehat{B}_t\widehat{x}_{t-1} \ (t = 2, ..., T),$$

and

$$\widehat{e}_T := \widehat{c}_T \ , \ \ \widehat{e}_t := \widehat{c}_t - \mathbb{E}[\widehat{B}_{t+1}^\top \widehat{u}_{t+1} \,|\, \widehat{\mathcal{F}}_t] \ (t = 2, ..., T).$$

For any $t \in \{1, ..., T\}$ and $\omega \in \Omega$, we analyze the primal-dual LP (in the sense of Section 2.1)

$$\begin{aligned}&\min_{\mathtt{y}_t \in \mathbb{R}^{n_t}} \{\widehat{e}_t^\top(\omega)\mathtt{y}_t \mid A_t\mathtt{y}_t = \widehat{d}_t(\omega), \ \mathtt{y}_t \ge 0\} \\ = &\max_{\mathtt{v}_t \in \mathbb{R}^{m_t}, \mathtt{r}_t \in \mathbb{R}^{n_t}} \{\widehat{d}_t^\top(\omega)\mathtt{v}_t \mid A_t^\top \mathtt{v}_t + \mathtt{r}_t = \widehat{e}_t(\omega), \ \mathtt{r}_t \ge 0\}. \qquad (4.16)\end{aligned}$$

Note that the aggregated solutions $\widehat{x}_t(\omega)$ and $(\widehat{u}_t(\omega), \widehat{s}_t(\omega))$ are almost surely feasible in the primal-dual pair (4.16). They are even almost surely optimal because of the complementarity $\widehat{s}_t^\top \widehat{x}_t = 0$ (a.s.), $(t = 1, ..., T)$. Furthermore, by Definition 4.10 one has $\widehat{d}_1 = b_1 \in \text{int}(A_1\mathbb{R}_+^{n_1})$, whereas $\widehat{d}_t \in \text{int}(A_t\mathbb{R}_+^{n_t})$

(a.s.) $(t = 2, ..., T)$ holds by (4.13). Now we can apply Proposition 2.1 e), and hence, there are constants $\big(L_t = L_t(A_t)\big)_{t=1,...,T}$ such that

$$\left|\begin{pmatrix}\widehat{u}_t\\ \widehat{s}_t\end{pmatrix}\right| \le L_t|\widehat{e}_t| \text{ (a.s.) } (t = 1, ..., T).$$

We verify the existence of model constants $\gamma_1, ..., \gamma_T \in \mathbb{R}_+$, independent of $\widehat{\mathcal{F}}$, such that

$$\left|\begin{pmatrix}\widehat{u}_t\\ \widehat{s}_t\end{pmatrix}\right| \le \gamma_t \text{ (a.s.) },\ (t = 1, ..., T). \tag{4.17}$$

It holds

$$\left|\begin{pmatrix}\widehat{u}_T\\ \widehat{s}_T\end{pmatrix}\right| \le L_T|\widehat{e}_T| = L_T|\widehat{c}_T| \le L_T\text{ess sup}|\widehat{c}_T| \le L_T\text{ess sup}|c_T| =: \gamma_T,\ \text{(a.s.)},$$

and, inductively for $t = T-1, ..., 1$,

$$\begin{aligned}\left|\begin{pmatrix}\widehat{u}_t\\ \widehat{s}_t\end{pmatrix}\right| &\le L_t|\widehat{e}_t| = L_t\Big|\widehat{c}_t - \mathbb{E}[\widehat{B}_{t+1}^\top\widehat{u}_{t+1}\,|\,\widehat{\mathcal{F}}_t]\Big|\\ &\le L_t\Big(|\widehat{c}_t| + \mathbb{E}\big[|\widehat{B}_{t+1}^\top||\widehat{u}_{t+1}|\,\big|\,\widehat{\mathcal{F}}_t\big]\Big)\\ &\le L_t\Big(\text{ess sup}|\widehat{c}_t| + \text{ess sup}\big(|\widehat{B}_{t+1}^\top||\widehat{u}_{t+1}|\big)\Big)\\ &\le L_t\big(\text{ess sup}|c_t| + \text{ess sup}|B_{t+1}^\top|\gamma_{t+1}\big) =: \gamma_t,\ \text{(a.s.)}.\end{aligned}$$

Thus (4.17) is verified, and the assertion of the lemma follows since

$$\left|\begin{pmatrix}\widehat{u}\\ \widehat{s}\end{pmatrix}\right| \le \textstyle\sum_{t=1}^T\left|\begin{pmatrix}\widehat{u}_t\\ \widehat{s}_t\end{pmatrix}\right| \le \sum_{t=1}^T\gamma_t =: \gamma \text{ (a.s.)}. \qquad \square$$

The result of combining Theorem 3.14, Theorem 4.7, Lemma 4.8 and Lemma 4.11 is stated in

Theorem 4.12
Assume that

W1) $\mathcal{F}_t = \sigma(\xi_2, ..., \xi_t)$ $(t = 2, ..., T)$ *where* $\xi_2, ..., \xi_T$ *are independent random vectors;*

W2) - A_t, c_t *are fixed,*
- $B_t \in L^\infty\big(\sigma(\xi_t); \mathbb{R}^{m_t\times n_{t-1}}\big)$, $b_t \in L^\infty\big(\sigma(\xi_2, ..., \xi_t); \mathbb{R}^{m_t}\big)$, $(t = 2, ..., T)$;

W3) $\{\mathsf{z}_t \in \mathbb{R}^{n_t} \mid A_t\mathsf{z}_t = 0,\ \mathsf{z}_t \ge 0\} = \{0\}$, $(t = 1, ..., T)$;

W4) `MSLP-`$\mathcal{P}(\mathcal{F})$ *is RCR*^{o} *and supp*$\big\{(B_2, b_2), ..., (B_T, b_T)\big\}$ *is convex.*

For a given finitely generated subfiltration $\widehat{\mathcal{F}}$ *of* $\mathcal{F}$*, having that*

$$\widehat{\mathcal{F}}_t \subset \sigma(\widehat{\mathcal{F}}_{t-1}, \xi_t) \ \ (t = 2, ..., T),$$

let $\widehat{x} \in$ *argmin*$\big($`MSLP-`$\mathcal{P}(\widehat{\mathcal{F}})\big)$, $(\widehat{u}, \widehat{s}) \in$ *argmax*$\big($`MSLP-`$\mathcal{D}(\widehat{\mathcal{F}})\big)$ *and* $\overline{x}$ *be recursively defined as in Theorem 4.7. Then* $\overline{x}$ *and* $(\widehat{u}, \widehat{s})$ *are feasible in* `MSLP-`$\mathcal{P}(\mathcal{F})$ *and* `MSLP-`$\mathcal{D}(\mathcal{F})$*, respectively, and their complementarity variable is bounded by*

$$0 \leq \widehat{s}^\top \overline{x} \leq C {\textstyle\sum_{t=2}^T} \Big(|b_t - \widehat{b}_t| + |B_t - \widehat{B}_t|\Big) \ \text{(a.s.)} \tag{4.18}$$

where $C \in \mathbb{R}_+$ *is a model constant (independent of* $\widehat{\mathcal{F}}$*). Moreover, the assumption W3) is superfluous whenever* $(B_t)_{t=2,..,T}$ *is fixed. In this case, (4.18) reads as*

$$0 \leq \widehat{s}^\top \overline{x} \leq C {\textstyle\sum_{t=2}^T} |b_t - \widehat{b}_t| \ \text{(a.s.)}. \tag{4.19}$$

Proof. By using W1) and W2), the feasibility of $(\widehat{u}, \widehat{s})$ in the original dual follows from Theorem 3.14. Because W4) implies particularly that `MSLP-`$\mathcal{P}(\mathcal{F})$ is RCR, one can apply Theorem 4.7. Thus, $\overline{x}$ is feasible in the original primal problem with

$$|\overline{x} - \widehat{x}| \leq L {\textstyle\sum_{t=2}^T} \Big(|b_t - \widehat{b}_t| + |B_t - \widehat{B}_t||\widehat{x}_{t-1}|\Big) \ \text{(a.s.)}$$

where $L \in \mathbb{R}_+$ is a model constant. In the case that $(B_t)_{t=2,...,T}$ is fixed, the relation $B_t = \widehat{B}_t$ $(t = 2, ..., T)$ holds true and one obtains $|\overline{x} - \widehat{x}| \leq L {\textstyle\sum_{t=2}^T} |b_t - \widehat{b}_t|$ (a.s.). Otherwise, assuming W3), Lemma 4.8 ensures that

$$|\overline{x} - \widehat{x}| \leq L \max\{1, \beta\} {\textstyle\sum_{t=2}^T} \Big(|b_t - \widehat{b}_t| + |B_t - \widehat{B}_t|\Big) \ \text{(a.s.)}.$$

Furthermore, because $\widehat{x}$ and $(\widehat{u}, \widehat{s})$ are supposed to be optimal in the aggregated (and finite discrete) primal-dual problem, it holds $\widehat{s}^\top \widehat{x} = 0$ (a.s.). Since also $\overline{x} \geq 0$ (a.s.), one has

$$0 \leq \widehat{s}^\top \overline{x} = \widehat{s}^\top \overline{x} - \widehat{s}^\top \widehat{x} \leq |\widehat{s}||\overline{x} - \widehat{x}| \leq \gamma |\overline{x} - \widehat{x}|, \ \text{(a.s.)},$$

where $\gamma \in \mathbb{R}_+$ is the model constant of Lemma 4.11 by using W4). Hence if $(B_t)_{t=2,...,T}$ is fixed, then let $C := \gamma L \in \mathbb{R}_+$, otherwise let $C := \gamma L \max\{1, \beta\} \in \mathbb{R}_+$. □

For the sake of completeness we also provide a slightly weaker version by

Theorem 4.13
Given the same setting as in Theorem 4.12 except that W4) is replaced with the weaker assumption
W4') $\texttt{MSLP-}\mathcal{P}(\mathcal{F})$ is RCR.

Then the same statements hold except that (4.18) and (4.19) have to be replaced with the weaker versions

$$0 \leq \mathbb{E}[\widehat{s}^\top \overline{x}] \leq C\mathbb{E}\left[\sum_{t=2}^T \left(|b_t - \widehat{b}_t| + |B_t - \widehat{B}_t|\right)\right] \tag{4.20}$$

and - if $(B_t)_{t=2,...,T}$ is fixed -

$$0 \leq \mathbb{E}[\widehat{s}^\top \overline{x}] \leq C\mathbb{E}\left[\sum_{t=2}^T |b_t - \widehat{b}_t|\right], \tag{4.21}$$

where $C \in \mathbb{R}_+$ is a model constant (independent of $\widehat{\mathcal{F}}$). Again, W3) is superfluous for the validity of (4.21).

Proof. Contrary to Lemma 4.11, W4') permits at least to apply Theorem 4.7. Hence $\overline{x}$ and $(\widehat{u}, \widehat{s})$ are feasible in $\texttt{MSLP-}\mathcal{P}(\mathcal{F})$ and $\texttt{MSLP-}\mathcal{D}(\mathcal{F})$, respectively, producing a duality gap of

$$\begin{aligned}
0 &\leq \mathbb{E}[\widehat{s}^\top \overline{x}] = \mathbb{E}[\mathbf{c}^\top \overline{x}] - \mathbb{E}[\mathbf{b}^\top \widehat{u}] = \mathbb{E}[\mathbf{c}^\top \overline{x}] - \max\left(\texttt{MSLP-}\mathcal{D}(\widehat{\mathcal{F}})\right) \\
&= \mathbb{E}[\mathbf{c}^\top \overline{x}] - \min\left(\texttt{MSLP-}\mathcal{P}(\widehat{\mathcal{F}})\right) = \mathbb{E}[\mathbf{c}^\top \overline{x}] - \mathbb{E}[\mathbf{c}^\top \widehat{x}] \\
&\leq \mathbb{E}\left[|\mathbf{c}||\overline{x} - \widehat{x}|\right] = |\mathbf{c}|\,\mathbb{E}|\overline{x} - \widehat{x}|,
\end{aligned}$$

noting that the cost vector $\mathbf{c} = (c_1, ..., c_T)^\top$ is fixed. According to the proof of Theorem 4.12, let $C := |\mathbf{c}|L \in \mathbb{R}_+$ when $(B_t)_{t=2,...,T}$ is fixed, otherwise let $C := |\mathbf{c}|L\max\{1, \beta\} \in \mathbb{R}_+$. □

The duality gap $\mathbb{E}[\widehat{s}^\top \overline{x}]$ can basically be quite small even if the subfiltration $\widehat{\mathcal{F}}$ is rather coarse, as demonstrated by some numerical tests in Chapter 6. However, the worst-case in terms of numerical aspects would be to generate an ascending chain $(\widehat{\mathcal{F}}^{(k)})_{k\geq 0}$ of subfiltrations of $\mathcal{F}$ with

$$\lim_{k\to\infty} \mathbb{E}\left[\sum_{t=2}^T \left(|b_t - \widehat{b}_t^{(k)}| + |B_t - \widehat{B}_t^{(k)}|\right)\right] = 0 \tag{4.22}$$

in order to ensure by (4.20) that $\lim_{k\to\infty} \mathbb{E}[\widehat{s}^{(k)\top} \overline{x}^{(k)}] = 0$. But in our strategy of local refinements, see next chapter, it is not necessary to assume the property (4.22) for the convergence proof. For this purpose, the setting of Theorem 4.12 will be needed rather than the weaker version of Theorem 4.13.

Example 4.14

We shall give an example where the 'expected tightness' (4.21) holds in contrast to the stronger version (4.19). Let us consider the three-stage problem

$$\texttt{MSLP-}\mathcal{P}(\mathcal{F}): \quad \textit{Minimize}_x \ \mathbb{E}\Big[x_1 + x_{2_1}(\xi_2) + x_{2_2}(\xi_2) + x_3(\xi_2,\xi_3)\Big]$$

$$\begin{array}{lllll} x_1 & & & = & 1 \\ x_1 & + & x_{2_1}(\xi_2) - x_{2_2}(\xi_2) & = & \xi_2 \quad \text{(a.s.)} \\ & & x_{2_1}(\xi_2) + x_{2_2}(\xi_2) \quad - \quad x_3(\xi_2,\xi_3) & = & \xi_3 \quad \text{(a.s.)} \\ & & x_1, x_2(\xi_2), x_3(\xi_2,\xi_3) & \geq & 0 \quad \text{(a.s.)} \end{array}$$

where ξ_2 and ξ_3 are stochastically independent, ξ_2 takes on the values -1 and 3 with equal probability and ξ_3 is uniformly distributed on $[0,1]$. The decisions x_1 and x_3 consist of one component only, whereas x_2 is built in (x_{2_1}, x_{2_2}). First we note that W1) and W2) of Theorem 4.12 are satisfied, whereas W3) is superfluous since the matrices $B_2 = 1$, $B_3 = (1,1)$ are fixed. Moreover, we show that MSLP-$\mathcal{P}(\mathcal{F})$ *has the RCR^o property (see Definition 4.10). Since $x_1 = 1$ is enforced, it follows that $x_{2_1}(\xi_2) - x_{2_2}(\xi_2) = \xi_2 - 1 \in \{-2,2\}$ (a.s.) for every feasible continuation $x_2(\xi_2)$ of $x_1 = 1$. Since also $x_{2_1}(\xi_2), x_{2_2}(\xi_2) \geq 0$ (a.s.), it holds $x_{2_1}(\xi_2) + x_{2_2}(\xi_2) \geq 2$ (a.s.), and therefore $x_3(\xi_2,\xi_3) = x_{2_1}(\xi_2) + x_{2_2}(\xi_2) - \xi_3 \geq 2 - \xi_3 \geq 1$ (a.s.). We conclude that* MSLP-$\mathcal{P}(\mathcal{F})$ *is RCR^o, and hence, it is also RCR. But contrary to Theorem 4.13, Theorem 4.12 is not applicable because supp$\{b_2\}$ = supp$\{\xi_2\} = \{-1,3\}$ is not convex.*

In Appendix A we construct an ascending chain of subfiltrations $(\widehat{\mathcal{F}}^{(k)})_{k\geq 0}$ of $\mathcal{F}$ having $\widehat{\mathcal{F}}_1^{(k)} = \widehat{\mathcal{F}}_2^{(k)} = \{\emptyset, \Omega\}$ and $\widehat{\mathcal{F}}_3^{(k)} \subset \widehat{\mathcal{F}}_3^{(k+1)} \subset \sigma(\xi_3)$, $\forall k \geq 0$. The example yields that

- MSLP-$\mathcal{P}(\widehat{\mathcal{F}}^{(k)})$ *has a unique solution $\widehat{x}^{(k)}$, $\forall k \geq 0$,*
- MSLP-$\mathcal{D}(\widehat{\mathcal{F}}^{(k)})$ *has a unique solution $(\widehat{u}^{(k)}, \widehat{s}^{(k)})$, $\forall k \geq 0$,*
- *the recursively defined policy $\overline{x}^{(k)}$ of Theorem 4.7 coincide with the unique optimal solution $\overline{x}$ of* MSLP-$\mathcal{P}(\mathcal{F})$, $\forall k \geq 0$,
- $\lim_{k\to\infty}$ *ess sup* $\left\{\widehat{s}^{(k)\top}\overline{x}^{(k)}\right\} = +\infty$, *and therefore (4.19) fails because the right side in (4.19) is uniformely essentially bounded.*

Remarks 4.15

An alternative and simpler policy than $\overline{x} = (\widehat{x}_1, \overline{x}_2, ..., \overline{x}_T)$ of Theorem 4.7 is obtained by

- $\overline{\overline{x}}_1 := \widehat{x}_1$,

- *for* $t = 2, ..., T$, $\forall \omega \in \Omega$,

$$\overline{\overline{x}}_t(\omega) := \operatorname*{argmin}_{\mathrm{y}_t \in \mathbb{R}^{n_t}} \left\{ |\mathrm{y}_t - \widehat{x}_t(\omega)| \;\middle|\; A_t \mathrm{y}_t = b_t(\omega) - B_t(\omega)\overline{\overline{x}}_{t-1}(\omega),\ \mathrm{y}_t \geq 0 \right\}$$

where the optimal dual slack $\widehat{s}$ *of* `MSLP-`$\mathcal{D}(\widehat{\mathcal{F}})$ *is not involved in this setting. One can prove a similar result for Theorem 4.7 and Theorem 4.12 where - unfortunately - the identical majorant on the right-hand side in (4.8) and (4.18), respectively, holds by using* $\overline{\overline{x}}$ *instead of* $\overline{x}$*. That is to say, some similar proofs do not detect a quantitative difference of the worst-case behavior between* $\widehat{s}^\top \overline{x}$ *and* $\widehat{s}^\top \overline{\overline{x}}$*. Note that in Theorem 4.7 the second stage complementarity part* $\widehat{s}_2^\top \overline{x}_2$ *is pointwise given by*

$$\widehat{s}_2(\omega)^\top \overline{x}_2(\omega) = \min_{\mathrm{y}_2 \in \mathbb{R}^{n_2}} \left\{ \widehat{s}_2(\omega)^\top \mathrm{y}_2 \;\middle|\; A_2 \mathrm{y}_2 = b_2(\omega) - B_2(\omega)\widehat{x}_1,\ \mathrm{y}_2 \geq 0 \right\}.$$

Therefore, in the two-stage case it obviously holds

$$0 \leq \widehat{s}^\top \overline{x} = \widehat{s}_1^\top \widehat{x}_1 + \widehat{s}_2^\top \overline{x}_2 = \widehat{s}_2^\top \overline{x}_2 \leq \widehat{s}_2^\top \overline{\overline{x}}_2 = \widehat{s}_1^\top \widehat{x}_1 + \widehat{s}_2^\top \overline{\overline{x}}_2 = \widehat{s}^\top \overline{\overline{x}}\ , \ \text{(a.s.)}.$$

Hence in the case $T = 2$*, one always obtains*

$$\inf \big(\texttt{MSLP-}\mathcal{P}(\mathcal{F})\big) \leq \mathbb{E}[\mathbf{c}^\top \overline{x}] = \mathbb{E}[\mathbf{b}^\top \widehat{u}] + \mathbb{E}[\widehat{s}^\top \overline{x}] \leq \mathbb{E}[\mathbf{b}^\top \widehat{u}] + \mathbb{E}[\widehat{s}^\top \overline{\overline{x}}] = \mathbb{E}[\mathbf{c}^\top \overline{\overline{x}}].$$

That is why we omit the simpler recursion $\overline{\overline{x}}$ *in the following.*

Open question:

There are probably some instances where $T \geq 3$ *and neither* $\widehat{s}^\top \overline{x}$ *is dominated by* $\widehat{s}^\top \overline{\overline{x}}$ *nor vice versa. Are there some individual majorants for* $\widehat{s}^\top \overline{x}$ *and* $\widehat{s}^\top \overline{\overline{x}}$ *in order to express the worst-case behavior in a different way? Which policy,* $\overline{x}$ *or* $\overline{\overline{x}}$*, results in a lower objective value for special data and distributions in the case of* $T \geq 3$*?*

5 Local refinements

The main subject of this chapter is a proposed approximation scheme for $\mathtt{MSLP}\text{-}\mathcal{P}(\mathcal{F})$, named `MSLP-APPROX`. It is based on simulated outcomes $\Delta(\omega) := \widehat{s}(\omega)^\top \overline{x}(\omega)$ of the previously investigated complementarity variable $\Delta = \widehat{s}^\top \overline{x} : \Omega \to [0, +\infty)$ in combination with local disaggregation of the associated subfiltration $\widehat{\mathcal{F}}$ where the expected value problem provides the initial $\widehat{\mathcal{F}}$. We shall obtain a probabilistic quality measure of a candidate solution $\overline{x} = (\overline{x}_1, ..., \overline{x}_T)$ at the stopping time of `MSLP-APPROX` which is shown to exist. Furthermore, an infinite extension will result in the method `MSLP-SOLVE` where it is proved that, by successively increasing the sample size and the accuracy parameter of `MSLP-APPROX`, the (weak) accumulation points of the candidate solutions solve the original problem.

During the chapter we assume M1)-M3):

M1) The information field at stage $t = 1, ..., T$ is given by the filtration $\mathcal{F} = \left(\mathcal{F}_t\right)_{t=1,...,T}$ where $\mathcal{F}_1 = \{\emptyset, \Omega\}$ and $\mathcal{F}_t = \sigma(\xi_2, ..., \xi_t)$ $(t = 2, ..., T)$ with $\xi_2, ..., \xi_T$ being stochastically independent and essentially bounded random vectors having the support

$$\operatorname{supp}\{\xi_t\} = [\![\alpha_t, \beta_t]\!] := \times_{i=1}^{l_t} [\alpha_{ti}, \beta_{ti}] \subset \mathbb{R}^{l_t}$$

where $\alpha_t, \beta_t \in \mathbb{R}^{l_t}$, $\alpha_t < \beta_t$, and - for technical reasons - $\mathbb{P}[\xi_{ti} = \beta_{ti}] = 0$ $(i = 1, ..., l_t)$, $(t = 2, ..., T)$; note that the components $(\xi_{ti})_{i=1,..,l_t}$ of ξ_t do not need to be independent.

M2) Given $A_1 \in \mathbb{R}^{\mathtt{m}_1 \times \mathtt{n}_1}$, $b_1 \in \mathbb{R}^{\mathtt{m}_1}$, $c_1 \in \mathbb{R}^{\mathtt{n}_1}$, and

$$A_t \in \mathbb{R}^{\mathtt{m}_t \times \mathtt{n}_t},\ B_t = \mathtt{B}_t(\xi_t),\ b_t = \mathtt{b}_t(\xi_2, ..., \xi_t),\ c_t \in \mathbb{R}^{\mathtt{n}_t} \quad (t = 2, ..., T),$$

where $\mathtt{B}_t : \mathbb{R}^{l_t} \longrightarrow \mathbb{R}^{\mathtt{m}_t \times \mathtt{n}_{t-1}}$ and $\mathtt{b}_t : \mathbb{R}^{l_2 + ... + l_t} \longrightarrow \mathbb{R}^{\mathtt{m}_t}$ are affine linear in their arguments as a whole.

M3) Suppose that

- $\mathtt{MSLP}\text{-}\mathcal{P}(\mathcal{F})$ is RCR^o (cf. Definition 4.10);
- $\mathtt{MSLP}\text{-}\mathcal{D}(\mathcal{T})$ is feasible where $\mathcal{T} := \{\emptyset, \Omega\}_{t=1,...,T}$ (cf. (3.13));
-

$$\{\mathtt{z}_t \in \mathbb{R}^{\mathtt{n}_t} \mid A_t \mathtt{z}_t = 0,\ \mathtt{z}_t \geq 0\} = \{0\} \ \ (t = 1, ..., T), \tag{5.1}$$

whereas (5.1) is superfluous in the case that $(B_t)_{t=2,...,T}$ is fixed.

Note that M1)-M3) permit to apply Theorem 3.14 and Theorem 4.12.

5.1 Composing and refining subfiltrations

As seen in Section 3.4, replacing the filtration $\mathcal{F}$ with any finitely generated subfiltration $\widehat{\mathcal{F}} \subset \mathcal{F}$ leads not to lower bounds in any case. However, due to M1)-M3) and Theorem 3.14 a lower bound to the unknown optimal objective is obtained if the subfiltration $\widehat{\mathcal{F}}$ satisfies

$$\widehat{\mathcal{F}}_t \subset \sigma(\widehat{\mathcal{F}}_{t-1}, \xi_t) \ \ (t = 2, ..., T). \tag{5.2}$$

In Remarks 3.12 1), a counterexample to (5.2) has been mentioned due to Wright [30] by taking $T = 3$ and $\{\emptyset, \Omega\} = \widehat{\mathcal{F}}_1 = \widehat{\mathcal{F}}_2 \subsetneqq \widehat{\mathcal{F}}_3 \subset \sigma(\xi_2)$ what is in contradiction to $\widehat{\mathcal{F}}_3 \subset \sigma(\widehat{\mathcal{F}}_2, \xi_3) = \sigma(\xi_3)$.

We now use the nodal syntax of Section 3.1 to construct finitely generated subfiltrations $\widehat{\mathcal{F}}$ of $\mathcal{F}$ with the property (5.2). This means that each $\widehat{\mathcal{F}}$ corresponds to a node set $\mathcal{N} = \bigcup_{t=1}^{T} \mathcal{N}_t$ consisting of disjoint node sets $\mathcal{N}_t$ where $\mathcal{N}_1 = \{1\}$ is the root and, for $2 \leq t \leq T$, each node $n \in \mathcal{N}_t$ features a unique immediate ancestor node $p_n \in \mathcal{N}_{t-1}$ at stage $t-1$, whereas each node $m \in \mathcal{N}_{t-1}$ has a nonempty set of immediate successors $C_m := \{n \in \mathcal{N}_t \mid p_n = m\}$ (see also Notations 3.1). Note that by M1) there are vectors $\alpha_t, \beta_t \in \mathbb{R}^{l_t}$, $\alpha_t < \beta_t$, such that $\xi_t \in [\![\alpha_t, \beta_t)\!) = \times_{i=1}^{l_t} [\alpha_{ti}, \beta_{ti}) \subset \mathbb{R}^{l_t}$ (a.s.), $(t = 2, ..., T)$. The procedure of disaggregation will generate an ascending chain $(\widehat{\mathcal{F}}^{(k)})_{k\geq 0}$ of finitely generated subfiltrations of $\mathcal{F}$, and therefore, the chain is associated with a sequence of trees $(\mathcal{N}^{(k)})_{k\geq 0}$. The algorithmic description of successive disaggregation looks as follows:

(i) Set $k := 0$ and begin with the trivial tree $\mathcal{N}^{(k)}$ belonging to singletons $\mathcal{N}_t^{(k)} := \{t\}$ $(t = 1, ..., T)$. Let $\alpha_t^{[n]} := \alpha_t$, $\beta_t^{[n]} := \beta_t$, $n \in \mathcal{N}_t^{(k)}$ $(t = 2, ..., T)$.

(ii) Define $\Omega^{[1]} := \Omega$ and $\Omega^{[n]} := \Omega^{[p_n]} \cap \xi_t^{-1}\big\{[\![\alpha_t^{[n]}, \beta_t^{[n]})\!)\big\}$, $n \in \mathcal{N}_t^{(k)}$ $(t = 2, ..., T)$, and let $\widehat{\mathcal{F}}_t^{(k)} := \sigma\big((\Omega^{[n]})_{n\in\mathcal{N}_t^{(k)}}\big)$ $(t = 1, ..., T)$.

Select

$$1.)\ t \in \{2, ..., T\} \quad 2.)\ n \in \mathcal{N}_t^{(k)} \quad 3.)\ i \in \{1, ..., l_t\} \quad 4.)\ \gamma \in (\alpha_{ti}^{[n]}, \beta_{ti}^{[n]}). \tag{5.3}$$

Then the selected node n together with its branch of all successor nodes is duplicated (see also Fig. 5.1). This results in a larger node set $\mathcal{N}^{(k+1)}$. Suppose that n has been separated into $n', n'' \in \mathcal{N}_t^{(k+1)}$. According to

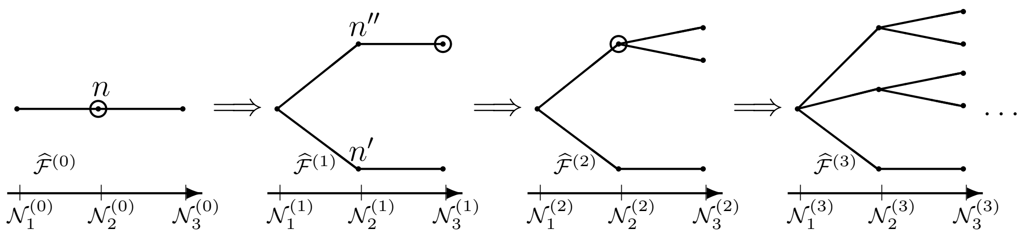

Figure 5.1: *Successive tree splitting associated with an ascending chain of subfiltrations of $\mathcal{F}$. Starting from $\widehat{\mathcal{F}}^{(0)} = \{\emptyset, \Omega\}_{t=1,..,T}$, a node is labeled for splitting.*

(5.3), the new rectangles assigned to n' and n'' are given by

$$\begin{aligned}[\![\alpha_t^{[n']}, \beta_t^{[n']})\!) \quad := \quad & [\alpha_{t1}^{[n]}, \beta_{t1}^{[n]}) \times \cdots \times [\alpha_{t\,i-1}^{[n]}, \beta_{t\,i-1}^{[n]}) \\ & \times [\alpha_{ti}^{[n]}, \gamma) \times [\alpha_{t\,i+1}^{[n]}, \beta_{t\,i+1}^{[n]}) \times \cdots \times [\alpha_{t\,l_t}^{[n]}, \beta_{t\,l_t}^{[n]})\end{aligned}$$

and

$$\begin{aligned}[\![\alpha_t^{[n'']}, \beta_t^{[n'']})\!) \quad := \quad & [\alpha_{t1}^{[n]}, \beta_{t1}^{[n]}) \times \cdots \times [\alpha_{t\,i-1}^{[n]}, \beta_{t\,i-1}^{[n]}) \\ & \times [\gamma, \beta_{ti}^{[n]}) \times [\alpha_{t\,i+1}^{[n]}, \beta_{t\,i+1}^{[n]}) \times \cdots \times [\alpha_{t\,l_t}^{[n]}, \beta_{t\,l_t}^{[n]}).\end{aligned}$$

Each other node $m' \in \mathcal{N}_s^{(k+1)}$ inherits the rectangle $[\![\alpha_s^{[m]}, \beta_s^{[m]})\!)$ from its previous node $m \in \mathcal{N}_s^{(k)}$. Set $k := k+1$ and repeat (ii).

Lemma 5.1
The above procedure of successive disaggregation leads to an ascending chain $(\widehat{\mathcal{F}}^{(k)})_{k\geq 0}$ of subfiltrations of $\mathcal{F}$ having that

$$\widehat{\mathcal{F}}_t^{(k)} \subset \sigma(\widehat{\mathcal{F}}_{t-1}^{(k)}, \xi_t) \;\; (t = 2, ..., T), \; k = 0, 1, 2, .. . \tag{5.4}$$

Furthermore, for the rectangles $[\![\alpha_t^{[n]}, \beta_t^{[n]})\!) \subset \mathbb{R}^{l_t}$ associated with the nodes one has

$$\min_{2\leq t\leq T} \min_{n\in\mathcal{N}_t^{(k)}} \min_{1\leq i\leq l_t} \frac{\beta_{ti}^{[n]} - \alpha_{ti}^{[n]}}{\beta_{ti} - \alpha_{ti}} \leq \left(\frac{1}{k+1}\right)^{\frac{1}{l_2+\ldots+l_T}}, \; k = 0, 1, 2, .. . \tag{5.5}$$

Proof. By an induction argument on $k = 0, 1, 2...$, one easily verifies that each node $m \in \mathcal{N}_{t-1}^{(k)}$ at stage $t-1$ features an individual partition of the support of ξ_t (excluding β_t, but see M1)). This means particularly that

$$[\![\alpha_t, \beta_t)\!) = \bigcup_{n \in C_m} [\![\alpha_t^{[n]}, \beta_t^{[n]})\!), \ \forall m \in \mathcal{N}_{t-1}^{(k)} \ (t = 2, ..., T),$$

where the rectangles in each union are disjoint. The relation (5.4) is therefore an immediate consequence of the definition of the events $\Omega^{[n]} \in \mathcal{A}$ in (ii). Each node $n \in \mathcal{N}_t^{(k)}$, $t \geq 2$, has a unique immediate ancestor node at stage $t-1$, and hence, each leaf node $n \in \mathcal{N}_T^{(k)}$ has a unique ancestor node $n_t \in \mathcal{N}_t^{(k)}$ at stage $t = T-1, ..., 2$ where we let $n_T := n$. It follows that

$$\times_{t=2}^{T} [\![\alpha_t, \beta_t)\!) = \bigcup_{n \in \mathcal{N}_T^{(k)}} \times_{t=2}^{T} [\![\alpha_t^{[n_t]}, \beta_t^{[n_t]})\!) \quad \left(\subset \mathbb{R}^{l_2 + ... + l_T} \right) \tag{5.6}$$

where the sets in the union are disjoint. Hence, by changing over to the Lebesgue volume on both sides, one obtains

$$\prod_{t=2}^{T} \prod_{i=1}^{l_t} (\beta_{ti} - \alpha_{ti}) = \sum_{n \in \mathcal{N}_T^{(k)}} \prod_{t=2}^{T} \prod_{i=1}^{l_t} (\beta_{ti}^{[n_t]} - \alpha_{ti}^{[n_t]}) \ . \tag{5.7}$$

In each pass k, a single node n together with the subtree of all its successor nodes is duplicated. Thus, noting that $|\mathcal{N}_T^{(0)}| = 1$, it certainly holds $|\mathcal{N}_T^{(k)}| \geq k+1$, $\forall k \geq 0$, and equality (5.7) implies that

$$\begin{aligned}
1 &= \sum_{n \in \mathcal{N}_T^{(k)}} \prod_{t=2}^{T} \prod_{i=1}^{l_t} \frac{\beta_{ti}^{[n_t]} - \alpha_{ti}^{[n_t]}}{\beta_{ti} - \alpha_{ti}} \geq |\mathcal{N}_T^{(k)}| \min_{n \in \mathcal{N}_T^{(k)}} \prod_{t=2}^{T} \prod_{i=1}^{l_t} \frac{\beta_{ti}^{[n_t]} - \alpha_{ti}^{[n_t]}}{\beta_{ti} - \alpha_{ti}} \\
&\geq (k+1) \min_{n \in \mathcal{N}_T^{(k)}} \prod_{t=2}^{T} \prod_{i=1}^{l_t} \frac{\beta_{ti}^{[n_t]} - \alpha_{ti}^{[n_t]}}{\beta_{ti} - \alpha_{ti}} \\
&\geq (k+1) \min_{n \in \mathcal{N}_T^{(k)}} \left(\min_{2 \leq t \leq T} \min_{1 \leq i \leq l_t} \frac{\beta_{ti}^{[n_t]} - \alpha_{ti}^{[n_t]}}{\beta_{ti} - \alpha_{ti}} \right)^{l_2 + ... + l_T} \\
&= (k+1) \Big(\min_{n \in \mathcal{N}_T^{(k)}} \min_{2 \leq t \leq T} \min_{1 \leq i \leq l_t} \frac{\beta_{ti}^{[n_t]} - \alpha_{ti}^{[n_t]}}{\beta_{ti} - \alpha_{ti}} \Big)^{l_2 + ... + l_T} \\
&= (k+1) \Big(\min_{2 \leq t \leq T} \min_{n \in \mathcal{N}_t^{(k)}} \min_{1 \leq i \leq l_t} \frac{\beta_{ti}^{[n]} - \alpha_{ti}^{[n]}}{\beta_{ti} - \alpha_{ti}} \Big)^{l_2 + ... + l_T} .
\end{aligned}$$

This completes the proof of (5.5). □

5.2 A finite approximation scheme

We prove first that the aggregated data $(\widehat{B}_t, \widehat{b}_t)_{t=2,...,T}$ are determined by aggregating the random vectors $(\xi_t)_{t=2,...,T}$.

Lemma 5.2
Let $\widehat{\mathcal{F}}$ be a subfiltration of $\mathcal{F}$ with $\widehat{\mathcal{F}}_t \subset \sigma(\widehat{\mathcal{F}}_{t-1}, \xi_t)$ and let the conditional expectation of ξ_t with respect to $\widehat{\mathcal{F}}_t$ be denoted by $\widehat{\xi}_t := \mathbb{E}[\xi_t \mid \widehat{\mathcal{F}}_t]$, $(t = 2, ..., T)$. Then M1)-M2) imply that

$$\widehat{B}_t = \mathtt{B}_t(\widehat{\xi}_t),\ \widehat{b}_t = \mathtt{b}_t(\widehat{\xi}_2, ..., \widehat{\xi}_t) \quad (t = 2, ..., T). \tag{5.8}$$

Proof. Let $t \in \{2, ..., T\}$. By the affine linearity of $\mathtt{B}_t$ and $\mathtt{b}_t$ one has

$$\widehat{B}_t = \mathbb{E}[B_t \,|\, \widehat{\mathcal{F}}_t] = \mathbb{E}\big[\mathtt{B}_t(\xi_t) \,\big|\, \widehat{\mathcal{F}}_t\big] = \mathtt{B}_t\big(\mathbb{E}[\xi_t \mid \widehat{\mathcal{F}}_t]\big) = \mathtt{B}_t(\widehat{\xi}_t)$$

and

$$\widehat{b}_t = \mathbb{E}[b_t \,|\, \widehat{\mathcal{F}}_t] = \mathbb{E}\big[\mathtt{b}_t(\xi_2, ..., \xi_t) \,\big|\, \widehat{\mathcal{F}}_t\big] = \mathtt{b}_t\big(\mathbb{E}[\xi_2 \,|\, \widehat{\mathcal{F}}_t], ..., \mathbb{E}[\xi_t \,|\, \widehat{\mathcal{F}}_t]\big) \ .$$

Let $s \in \{2, ..., t\}$. The assumption on $\widehat{\mathcal{F}}$ implies particularly $\widehat{\mathcal{F}}_t \subset \sigma(\widehat{\mathcal{F}}_s, \xi_{s+1}, ..., \xi_t)$, and Lemma 3.7 a) states that $\mathbb{E}[\xi_s \,|\, \sigma(\widehat{\mathcal{F}}_s, \xi_{s+1}, ..., \xi_t)] = \mathbb{E}[\xi_s \,|\, \widehat{\mathcal{F}}_s]$. Hence it follows that

$$\begin{aligned} \mathbb{E}[\xi_s \,|\, \widehat{\mathcal{F}}_t] &= \mathbb{E}\big[\mathbb{E}[\xi_s \,|\, \sigma(\widehat{\mathcal{F}}_s, \xi_{s+1}, ..., \xi_t)] \,\big|\, \widehat{\mathcal{F}}_t\big] = \mathbb{E}\big[\mathbb{E}[\xi_s \,|\, \widehat{\mathcal{F}}_s] \,\big|\, \widehat{\mathcal{F}}_t\big] \\ &= \mathbb{E}[\xi_s \,|\, \widehat{\mathcal{F}}_s] = \widehat{\xi}_s. \end{aligned}$$

This completes the proof. □

An ascending chain of subfiltrations $(\widehat{\mathcal{F}}^{(k)})_{k\geq 0}$ generated as in the previous section corresponds to a sequence of approximating primal-dual problems $\mathtt{MSLP}\text{-}\mathcal{P}(\widehat{\mathcal{F}}^{(k)})$ and $\mathtt{MSLP}\text{-}\mathcal{D}(\widehat{\mathcal{F}}^{(k)})$, $k = 0, 1, 2, ...$ These problems can be formulated and solved as the finite dimensional LP‘s of Section 3.1, cf. (3.1) and (3.2). Therefore, we use the notations of the associated sequence of trees $(\mathcal{N}^{(k)})_{k\geq 0}$ including the rectangles $[\![\alpha_t^{[n]}, \beta_t^{[n]})\!) \subset \mathbb{R}^{l_t}$. The conditional expectation of ξ_t within each rectangles is denoted by

$$\overline{\xi}_t^{[n]} := \mathbb{E}\big[\xi_t \,\big|\, \xi_t \in [\![\alpha_t^{[n]}, \beta_t^{[n]})\!)\big],\ n \in \mathcal{N}_t^{(k)}\ (t = 2, ..., T). \tag{5.9}$$

Then, due to Lemma 5.2, the realizations of $\widehat{B}_t^{(k)}$ and $\widehat{b}_t^{(k)}$ at node $n \in \mathcal{N}_t^{(k)}$ are given by

$$B_t^{[n]} := \mathtt{B}_t\big(\overline{\xi}_t^{[n]}\big),\ b_t^{[n]} := \mathtt{b}_t\big(\overline{\xi}_2^{[n_2]}, ..., \overline{\xi}_t^{[n_t]}\big)$$

where $n_t := n$ and $n_s := p_{n_{s+1}} \in \mathcal{N}_s^{(k)}$ $(s = t-1, ..., 2)$, $(t = 2, ..., T)$, $k \geq 0$. Furthermore, let the probability of reaching node $n \in \mathcal{N}_t^{(k)}$ be denoted by

$$q^{[n]} := \mathbb{P}[\Omega^{[n]}] = \prod_{s=2}^{t} \mathbb{P}\left[\xi_s \in [\![\alpha_s^{[n_s]}, \beta_s^{[n_s]})\!)\right].$$

The aggregated primal and dual problems belonging to $\widehat{\mathcal{F}}^{(k)}$ can now be formulated as the LP's

$$\texttt{MSLP-}\mathcal{P}(\widehat{\mathcal{F}}^{(k)}): \qquad \min_x \quad \sum_{t=1}^{T} \sum_{n\in\mathcal{N}_t^{(k)}} q^{[n]} c_t^\top x_t^{[n]} \tag{5.10}$$

$$\text{subject to} \begin{cases} A_1 x_1^{[n]} &= b_1 \\ B_t^{[n]} x_{t-1}^{[p_n]} + A_t x_t^{[n]} &= b_t^{[n]} \quad, n \in \mathcal{N}_t^{(k)} \ (t = 2, ..., T) \\ x_t^{[n]} &\geq 0 \quad\ \ , n \in \mathcal{N}_t^{(k)} \ (t = 1, ..., T) \end{cases}$$

and

$$\texttt{MSLP-}\mathcal{D}(\widehat{\mathcal{F}}^{(k)}): \qquad \max_{u,s} \quad \sum_{t=1}^{T} \sum_{n\in\mathcal{N}_t^{(k)}} q^{[n]} b_t^{[n]\top} u_t^{[n]} \tag{5.11}$$

$$\text{subject to} \begin{cases} A_t^\top u_t^{[n]} + \sum_{m\in C_n} \frac{q^{[m]}}{q^{[n]}} B_{t+1}^{[m]\top} u_{t+1}^{[m]} + s_t^{[n]} &= c_t \ , n \in \mathcal{N}_t^{(k)} \\ & \qquad (t = 1, ..., T-1) \\ A_T^\top u_T^{[n]} + s_T^{[n]} &= c_T \ , n \in \mathcal{N}_T^{(k)} \\ s_t^{[n]} &\geq 0 \quad , n \in \mathcal{N}_t^{(k)} \\ & \qquad (t = 1, ..., T) \end{cases} .$$

Let $k \geq 0$ and assume that $\widehat{x}^{(k)} = (\widehat{x}_t^{[n]})_{n\in\mathcal{N}_t^{(k)}, t=1,...,T}$ and $(\widehat{u}^{(k)}, \widehat{s}^{(k)}) = (\widehat{u}_t^{[n]}, \widehat{s}_t^{[n]})_{n\in\mathcal{N}_t^{(k)}, t=1,...,T}$ solve $\texttt{MSLP-}\mathcal{P}(\widehat{\mathcal{F}}^{(k)})$ and $\texttt{MSLP-}\mathcal{D}(\widehat{\mathcal{F}}^{(k)})$, respectively. In Theorem 4.7, a recursive policy $\overline{x}^{(k)} = (\overline{x}_1^{(k)}, ..., \overline{x}_T^{(k)})$ in $\texttt{MSLP-}\mathcal{P}(\mathcal{F})$ has been obtained by $\widehat{x}^{(k)}$ and $(\widehat{u}^{(k)}, \widehat{s}^{(k)})$. We shall explain how to simulate a realization of the complementarity variable $\Delta^{(k)} := \widehat{s}^{(k)\top}\overline{x}^{(k)} : \Omega \longrightarrow [0, +\infty)$. According to (5.6), there is a unique assignment $\times_{t=2}^{T} [\![\alpha_t, \beta_t)\!) \to \times_{t=2}^{T} \mathcal{N}_t^{(k)}$ in the sense that almost every realized scenario belongs to a unique node path,

$$(\overline{\xi}_2, ..., \overline{\xi}_T) \longmapsto (\overline{n}_2, ..., \overline{n}_T), \tag{5.12}$$

where $\overline{\xi}_t \in [\![\alpha_t^{[\overline{n}_t]}, \beta_t^{[\overline{n}_t]})\!)$ and $\overline{n}_{t-1}$ is the immediate ancestor node of $\overline{n}_t$, $(t = 2, ..., T)$. After having sampled a scenario $(\overline{\xi}_2, ..., \overline{\xi}_T) \in \times_{t=2}^{T} [\![\alpha_t, \beta_t)\!)$ from the distribution of $(\xi_2, ..., \xi_T)$, let $\overline{x}_1 := \widehat{x}_1^{[1]}$ and compute recursively for $t = 2, ..., T$

$$\Delta_t^{(k)}(\overline{\xi}_2, ..., \overline{\xi}_t) := \min_z \left\{ \widehat{s}_t^{[\overline{n}_t]\top} z \;\middle|\; A_t z = \mathtt{b}_t(\overline{\xi}_2, ..., \overline{\xi}_t) - \mathtt{B}_t(\overline{\xi}_t)\overline{x}_{t-1},\ z \geq 0 \right\} \tag{5.13}$$

together with the solution of the quadratic problem

$$\overline{x}_t := \operatorname{argmin}_z \left\{ |z - \widehat{x}_t^{[\overline{n}_t]}|^2 \;\middle|\; \begin{array}{rcl} A_t z & = & \mathtt{b}_t(\overline{\xi}_2, ..., \overline{\xi}_t) - \mathtt{B}_t(\overline{\xi}_t)\overline{x}_{t-1}, \\ \widehat{s}_t^{[\overline{n}_t]\top} z & = & \Delta_t^{(k)}(\overline{\xi}_2, ..., \overline{\xi}_t), \\ z & \geq & 0 \end{array} \right\}, \tag{5.14}$$

whereas the latter problem is superfluous for $t = T$. Then the cumulative value

$$\Delta^{(k)}(\overline{\xi}_2, ..., \overline{\xi}_T) := \sum_{t=2}^{T} \Delta_t^{(k)}(\overline{\xi}_2, ..., \overline{\xi}_t) \tag{5.15}$$

is a sampled data of $\widehat{s}^{(k)\top}\overline{x}^{(k)} \ :\ \Omega \longrightarrow [0, +\infty)$ in the sense that $\Delta^{(k)}\big(\xi_2(\omega), ..., \xi_T(\omega)\big) = \widehat{s}^{(k)}(\omega)^\top \overline{x}^{(k)}(\omega)$ for almost all $\omega \in \Omega$.

We make three important remarks about $\Delta^{(k)} : \times_{t=2}^{T} [\![\alpha_t, \beta_t)\!) \to [0, +\infty)$:

- $\Delta^{(k)}$ is not uniquely determined by $\widehat{\mathcal{F}}^{(k)}$ because the optimal solutions of MSLP-$\mathcal{P}(\widehat{\mathcal{F}}^{(k)})$ or MSLP-$\mathcal{D}(\widehat{\mathcal{F}}^{(k)})$ do not need to be unique;
- $\Delta^{(k)}$ is generally not additive separable in its arguments $(\xi_2, ..., \xi_T) = (\xi_{21}, \xi_{22}, ..., \xi_{Tl_T})$;
- after a refinement $\widehat{\mathcal{F}}^{(k+1)} \supset \widehat{\mathcal{F}}^{(k)}$, the resulting new function $\Delta^{(k+1)}$ does not need to be pointwise smaller than the previous $\Delta^{(k)}$; in general, the monotonicity $\mathbb{E}\Delta^{(k+1)} \leq \mathbb{E}\Delta^{(k)}$ does not hold.

At least, one may apply Theorem 4.12 due to the assumptions M1)-M3) of page 53. Therefore, it follows that there is a model constant $C \in \mathbb{R}_+$, independent of k, such that

$$\begin{array}{rcl} 0 & \leq & \Delta^{(k)}(\xi_2, ..., \xi_T) \leq C\sum_{t=2}^{T} \left(|b_t - \widehat{b}_t^{(k)}| + |B_t - \widehat{B}_t^{(k)}| \right) \\ & \overset{(5.8)}{=} & C\sum_{t=2}^{T} \left(\left|\mathtt{b}_t(\xi_2, ..., \xi_t) - \mathtt{b}_t(\widehat{\xi}_2^{(k)}, ..., \widehat{\xi}_t^{(k)})\right| + \left|\mathtt{B}_t(\xi_t) - \mathtt{B}_t(\widehat{\xi}_t^{(k)})\right| \right) \\ & \leq & \overline{C}\sum_{t=2}^{T} |\xi_t - \widehat{\xi}_t^{(k)}| \leq \overline{C}\sum_{t=2}^{T}\sum_{i=1}^{l_t} |\xi_{ti} - \widehat{\xi}_{ti}^{(k)}| \ ,\ \text{(a.s.)}, \end{array} \tag{5.16}$$

where $\overline{C} \in \mathbb{R}_+$ depends on both C and the regression terms of the affine linear mappings $\big(\mathtt{B}_t(\cdot), \mathtt{b}_t(\cdot)\big)_{t=2,...,T}$. In other words, if $(\overline{\xi}_2, ..., \overline{\xi}_T) = (\overline{\xi}_{21}, ..., \overline{\xi}_{Tl_T})$ is a sampled scenario assigned to the node path $(\overline{n}_2, ..., \overline{n}_T)$ by (5.12), then it follows w.p.1 that

$$\Delta^{(k)}(\overline{\xi}_2, ..., \overline{\xi}_T) \leq \overline{C} \textstyle\sum_{t=2}^{T} \sum_{i=1}^{l_t} |\overline{\xi}_{ti} - \overline{\xi}_{ti}^{[\overline{n}_t]}| \tag{5.17}$$

where $\overline{\xi}_{ti}^{[\overline{n}_t]}$ is the i-th component of the conditional expected vector $\overline{\xi}_t^{[\overline{n}_t]}$ in (5.9). The goal is now to obtain suitable splitting rules by (5.3) to make $\Delta^{(k)}(\xi_2, ..., \xi_T)$ successively smaller in a certain probabilistic sense.

<u>`MSLP-APPROX`</u>

`Step 0`: *(Parameters and Initialization)*

Choose

- *a tolerance* $\varepsilon \in (0, +\infty)$,
- *an initial sample size* $N \in \mathbb{N}\backslash\{0\}$ *and an increment* $\varrho \in (0, +\infty)$,
- *scalars* $\{\lambda_{ti} > 0\}_{i=1}^{l_t}{}_{t=2}^{T}$.

Set $k := 0$ *and* $\widehat{\mathcal{F}}^{(k)} := \{\emptyset, \Omega\}_{t=1,...,T}$.

`Step 1`: *(Significance)*

Let $\Delta^{(k)} : \times_{t=2}^{T} [\![\alpha_t, \beta_t)\!) \longrightarrow [0, +\infty)$ *be the complementarity function associated with* $\widehat{\mathcal{F}}^{(k)}$ *and defined pointwise as in (5.15). Generate an i.i.d. sample* $(\overline{\xi}_2^{(j)}, ..., \overline{\xi}_T^{(j)})_{j=1,...,N_k}$ *of size* $N_k := \lceil Ne^{\varrho k} \rceil$ *from the distribution of* $(\xi_2, ..., \xi_T)$*. If*

$$\frac{1}{N_k} \textstyle\sum_{j=1}^{N_k} \Delta^{(k)}(\overline{\xi}_2^{(j)}, ..., \overline{\xi}_T^{(j)}) \leq \varepsilon, \tag{5.18}$$

then STOP. Else go to `Step 2`.

`Step 2`: *(Refinement)*

- *Choose* $j \in \{1, ..., N_k\}$ *with* $\Delta^{(k)}(\overline{\xi}_2^{(j)}, ..., \overline{\xi}_T^{(j)}) > \varepsilon$.
- *Let* $(\overline{\xi}_2^{(j)}, ..., \overline{\xi}_T^{(j)})$ *be assigned to the node path* $(\overline{n}_2, ..., \overline{n}_T)$ *associated with the sequence of expected vectors* $(\overline{\xi}_2^{[\overline{n}_2]}, ..., \overline{\xi}_T^{[\overline{n}_T]})$ *(cf. (5.12) and (5.9)). Select*

$$(\overline{t}, \overline{i}) \in \operatorname*{argmax}_{2 \leq t \leq T,\ 1 \leq i \leq l_t} \left\{ \lambda_{ti} |\overline{\xi}_{ti}^{(j)} - \overline{\xi}_{ti}^{[\overline{n}_t]}| \right\} . \tag{5.19}$$

- *Refine $\widehat{\mathcal{F}}^{(k)}$ to $\widehat{\mathcal{F}}^{(k+1)}$ according to the selection in (5.3) by*

$$t := \bar{t},\ n := \overline{n}_{\bar{t}},\ i := \bar{i},\ \gamma := \frac{1}{2}(\alpha_{\bar{t}\,\bar{i}}^{[\overline{n}_{\bar{t}}]} + \beta_{\bar{t}\,\bar{i}}^{[\overline{n}_{\bar{t}}]})\ .$$

- *Set $k := k+1$ and go to `Step 1`.*

Later in this section, some rules for refining simultaneously several intervals in `Step 2` are proposed. A probabilistic inference on the solution quality at the stopping time is analyzed in the next section. As already mentioned, $\Delta^{(k)}$ is not additive separable in its arguments. The selection in (5.19) means that the component $\overline{\xi}_{\bar{t}\,\bar{i}}^{(j)}$ is suspected to take on the highest ratio of $\Delta^{(k)}(\overline{\xi}_2^{(j)}, ..., \overline{\xi}_T^{(j)})\,(> \varepsilon)$. It is beyond the scope of this work to investigate suitable heuristics for the selection of the scalar parameters $\{\lambda_{ti}\}_{i=1}^{l_t}\,{}_{t=2}^{T}$. At least, the default values $\lambda_{ti} := \frac{1}{\beta_{ti}-\alpha_{ti}}$ $(i = 1, ..., l_t)$ $(t = 2, ..., T)$ seem to be appropriate if one presumes that all components have a similar influence on $\Delta^{(k)}$ given the event $\Delta^{(k)}(\overline{\xi}_2^{(j)}, ..., \overline{\xi}_T^{(j)}) > \varepsilon$. It should be noted that `MSLP-APPROX` can basically stop in a manageable number of iterations, even in the first iteration $k = 0$. But this strongly depends on the problem instance and the tolerance ε. However, the worst-case behavior is stated in

Proposition 5.3
`MSLP-APPROX` stops in less than $\kappa\varepsilon^{-(l_2+...+l_T)}$ iterations (w.p.1), where $\kappa \in \mathbb{R}_+$ depends on both a model constant and the scalars $\{\lambda_{ti} > 0\}_{i=1}^{l_t}\,{}_{t=2}^{T}$. Hence, κ is independent of the parameters ε, N, ϱ and all the simulated scenarios in `Step 1`.

Proof. Let

$$\kappa := \left(2\overline{C}(l_2 + ... + l_T)\frac{\max_{2\le t\le T, 1\le i\le l_t}\left\{\lambda_{ti}(\beta_{ti} - \alpha_{ti})\right\}}{\min_{2\le t\le T, 1\le i\le l_t}\lambda_{ti}}\right)^{l_2+...+l_T} \tag{5.20}$$

where $\{\lambda_{ti} > 0\}_{i=1}^{l_t}\,{}_{t=2}^{T}$ are the scalars and $\overline{C}$ is the model constant in (5.17). Now suppose that `MSLP-APPROX` does not stop before iteration $\widehat{k} := \lceil \kappa\varepsilon^{-(l_2+...+l_T)}\rceil \in \mathbb{N}\backslash\{0\}$. Inequality (5.5) of Lemma 5.1 states that there exist $\bar{t} \in \{1, ..., T\}$, $\widehat{n}_{\bar{t}} \in \mathcal{N}_{\bar{t}}^{(\widehat{k})}$ and $\bar{i} \in \{1, ..., l_{\bar{t}}\}$, associated with the interval $[\alpha_{\bar{t}\,\bar{i}}^{[\widehat{n}_{\bar{t}}]}, \beta_{\bar{t}\,\bar{i}}^{[\widehat{n}_{\bar{t}}]}) \subset [\alpha_{\bar{t}\,\bar{i}}, \beta_{\bar{t}\,\bar{i}})$ having the relative length

$$\frac{\beta_{\bar{t}\,\bar{i}}^{[\widehat{n}_{\bar{t}}]} - \alpha_{\bar{t}\,\bar{i}}^{[\widehat{n}_{\bar{t}}]}}{\beta_{\bar{t}\,\bar{i}} - \alpha_{\bar{t}\,\bar{i}}} \le \left(\frac{1}{\widehat{k}+1}\right)^{\frac{1}{l_2+...+l_T}} \le \left(\frac{1}{\kappa\varepsilon^{-(l_2+...+l_T)}}\right)^{\frac{1}{l_2+...+l_T}} = \varepsilon\left(\frac{1}{\kappa}\right)^{\frac{1}{l_2+...+l_T}}\ . \tag{5.21}$$

Moreover, since $\widehat{k} \geq 1$, one can assume that $[\alpha_{\bar{t}\bar{i}}^{[\widehat{n}_{\bar{t}}]}, \beta_{\bar{t}\bar{i}}^{[\widehat{n}_{\bar{t}}]})$ results from the splitting of an interval $[\alpha_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]}, \beta_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]})$ at a previous iteration $\overline{k} < \widehat{k}$, therefore $\overline{n}_{\bar{t}} \in \mathcal{N}_{\bar{t}}^{(\overline{k})}$. In `Step 2`, the splitting position γ of $[\alpha_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]}, \beta_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]})$ is provided to be the center. It follows that $[\alpha_{\bar{t}\bar{i}}^{[\widehat{n}_{\bar{t}}]}, \beta_{\bar{t}\bar{i}}^{[\widehat{n}_{\bar{t}}]})$ is equal to either $[\alpha_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]}, \frac{\alpha_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]}+\beta_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]}}{2})$ or $[\frac{\alpha_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]}+\beta_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]}}{2}, \beta_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]})$. In both cases one has by (5.21) that

$$\frac{\beta_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]} - \alpha_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]}}{\beta_{\bar{t}\bar{i}} - \alpha_{\bar{t}\bar{i}}} = \frac{2(\beta_{\bar{t}\bar{i}}^{[\widehat{n}_{\bar{t}}]} - \alpha_{\bar{t}\bar{i}}^{[\widehat{n}_{\bar{t}}]})}{\beta_{\bar{t}\bar{i}} - \alpha_{\bar{t}\bar{i}}} \leq 2\varepsilon \left(\frac{1}{\kappa}\right)^{\frac{1}{l_2+\ldots+l_T}} . \tag{5.22}$$

Furthermore, in `Step 1` of that iteration $\overline{k}$ there is a sample scenario $(\overline{\xi}_2, ..., \overline{\xi}_T) \in \times_{t=2}^{T} [\![\alpha_t, \beta_t))$ satisfying $\Delta^{(\overline{k})}(\overline{\xi}_2, ..., \overline{\xi}_T) > \varepsilon$ and $\overline{\xi}_{\bar{t}\bar{i}} \in [\alpha_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]}, \beta_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]})$, because otherwise $[\alpha_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]}, \beta_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]})$ would not have been intended to split in `Step 2`. Let $(\overline{\xi}_2^{[\overline{n}_2]}, ..., \overline{\xi}_T^{[\overline{n}_T]})$ denote the sequence of the expected vectors associated with $(\overline{\xi}_2, ..., \overline{\xi}_T)$ (cf. (5.9) and (5.12)). It follows that

$$\begin{array}{rcl}
\varepsilon & < & \Delta^{(k)}(\overline{\xi}_2, ..., \overline{\xi}_T) \overset{(5.17),\,\text{w.p.1}}{\leq} \overline{C} \sum_{t=2}^{T} \sum_{i=1}^{l_t} |\overline{\xi}_{ti} - \overline{\xi}_{ti}^{[\overline{n}_t]}| \\
& \leq & \left(\dfrac{\overline{C}}{\min_{2\leq t\leq T, 1\leq i\leq l_t} \lambda_{ti}}\right) \sum_{t=2}^{T} \sum_{i=1}^{l_t} \lambda_{ti} |\overline{\xi}_{ti} - \overline{\xi}_{ti}^{[\overline{n}_t]}| \\
& \leq & \left(\dfrac{\overline{C}(l_2 + ... + l_T)}{\min_{2\leq t\leq T, 1\leq i\leq l_t} \lambda_{ti}}\right) \max\limits_{2\leq t\leq T, 1\leq i\leq l_t} \left\{\lambda_{ti} |\overline{\xi}_{ti} - \overline{\xi}_{ti}^{[\overline{n}_t]}|\right\} \\
& \overset{(5.19)}{=} & \left(\dfrac{\overline{C}(l_2 + ... + l_T)}{\min_{2\leq t\leq T, 1\leq i\leq l_t} \lambda_{ti}}\right) \lambda_{\bar{t}\bar{i}} |\overline{\xi}_{\bar{t}\bar{i}} - \overline{\xi}_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]}| .
\end{array}$$

Note that both $\overline{\xi}_{\bar{t}\bar{i}}$ and the conditional expectation $\overline{\xi}_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]}$ lie in the interval $[\alpha_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]}, \beta_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]})$, hence $|\overline{\xi}_{\bar{t}\bar{i}} - \overline{\xi}_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]}| \leq (\beta_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]} - \alpha_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]})$. One concludes that

$$\begin{array}{rcl}
\varepsilon & < & \left(\dfrac{\overline{C}(l_2 + ... + l_T)}{\min_{2\leq t\leq T, 1\leq i\leq l_t} \lambda_{ti}}\right) \lambda_{\bar{t}\bar{i}} (\beta_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]} - \alpha_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]}) \\
& = & \left(\overline{C}(l_2 + ... + l_T) \dfrac{\lambda_{\bar{t}\bar{i}}(\beta_{\bar{t}\bar{i}} - \alpha_{\bar{t}\bar{i}})}{\min_{2\leq t\leq T, 1\leq i\leq l_t} \lambda_{ti}}\right) \dfrac{\beta_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]} - \alpha_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]}}{\beta_{\bar{t}\bar{i}} - \alpha_{\bar{t}\bar{i}}} \\
& \leq & \left(\overline{C}(l_2 + ... + l_T) \dfrac{\max_{2\leq t\leq T, 1\leq i\leq l_t} \left\{\lambda_{ti}(\beta_{ti} - \alpha_{ti})\right\}}{\min_{2\leq t\leq T, 1\leq i\leq l_t} \lambda_{ti}}\right) \dfrac{\beta_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]} - \alpha_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]}}{\beta_{\bar{t}\bar{i}} - \alpha_{\bar{t}\bar{i}}} \\
& \overset{(5.20)}{=} & \dfrac{1}{2} \kappa^{\frac{1}{l_2+\ldots+l_T}} \dfrac{\beta_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]} - \alpha_{\bar{t}\bar{i}}^{[\overline{n}_{\bar{t}}]}}{\beta_{\bar{t}\bar{i}} - \alpha_{\bar{t}\bar{i}}} \overset{(5.22)}{\leq} \dfrac{1}{2} \kappa^{\frac{1}{l_2+\ldots+l_T}} \, 2\varepsilon \left(\dfrac{1}{\kappa}\right)^{\frac{1}{l_2+\ldots+l_T}} = \varepsilon .
\end{array}$$

This is a contradiction, therefore, `MSLP-APPROX` stops in less than $\widehat{k} = \lceil \kappa\varepsilon^{-(l_2+\ldots+l_T)}\rceil$ iterations. $\square$

Proposition 5.3 certainly does not indicate the 'true' efficiency of `MSLP-APPROX` because the proof is obtained solely from the worst-case behavior of $\Delta^{(k)}$ due to (5.17) in combination with the general assertion in (5.5). The worst-case in terms of numerical aspects leads finally to a rectangle structure where

$$\max_{2\le t\le T}\ \max_{n\in\mathcal{N}_t^{(k)}}\ \max_{1\le i\le l_t} \frac{\beta_{ti}^{[n]}-\alpha_{ti}^{[n]}}{\beta_{ti}-\alpha_{ti}}$$

turns out to be small. However, our goal has simply been to show that the information coming solely from sampled values of $\Delta^{(k)}$ is sufficient to prove the finiteness with respect to the stopping criterion. On the other hand, when making the tolerance ε successively smaller by a simultaneous increase of the sample size, the sequence of the duality gaps $\mathbb{E}\Delta^{(\overline{k})}$ at the stopping time $\overline{k}$ tends to zero as discussed in the next section. In other words, the complementarity variable contains enough information for suitable refinements. This is the reason why the approach is not based on a direct approximation of the distribution of $(\xi_2,...,\xi_T)$ because, roughly speaking, at most the rectangles affected by the set

$$\left\{(\overline{\xi}_2,...,\overline{\xi}_T)\in \times_{t=2}^{T}[\![\alpha_t,\beta_t)\!)\,\middle|\,\Delta^{(k)}(\overline{\xi}_2,...,\overline{\xi}_T)>\varepsilon\right\}$$

are considered for refinement.

Remarks 5.4
In the report of Fusek et al. [10], some other rules for refining the rectangles have been proposed without having a primal counterpart in `MSLP-`$\mathcal{P}(\mathcal{F})$ *to the aggregated dual solutions. It should be mentioned that the convergence has been shown to fail in general. We shall briefly outline the main idea of [10]. The attempt has been reduced to the case that* B_t *(together with* A_t, c_t*) is fixed and the right-hand side* b_t *must not depend on the past, i.e.,* $b_t = \mathtt{b}_t(\xi_t)$, $t=2,...,T$. *In contrast to (5.13), the information for refinements is taken from the LP objectives*

$$\delta_t^{(k)}(\overline{\xi}_t) := \min_z \left\{\widehat{s}_t^{[\overline{n}_t]\top} z \,\middle|\, A_t z = \mathtt{b}_t(\overline{\xi}_t) - \mathtt{B}_t\widehat{x}_{t-1}^{[p_{\overline{n}_t}]},\ z\ge 0\right\},\ \ \overline{\xi}_t\in[\![\alpha_t^{[\overline{n}_t]},\beta_t^{[\overline{n}_t]})\!). \tag{5.23}$$

The main difference between (5.13) and (5.23) is therefore that the ancestor decision from stage $t-1$ *is modeled here as the aggregated decision* *coming from the ancestor node $p_{\overline{n}_t}$ of $\overline{n}_t$. Let $\overline{\delta}_t^{(k)}(\overline{n}_t)$ denote the Edmundson-Madansky upper bound for $\mathbb{E}\big[\delta_t^{(k)}(\xi_t) \bigm| \xi_t \in [\![\alpha_t^{[\overline{n}_t]}, \beta_t^{[\overline{n}_t]})\!)\big]$, see also Kall/Mayer [16] or Section 3.3. The method suggests to refine all those rectangles where $\overline{\delta}_t^{(k)}(\overline{n}_t)$ is above a tolerance ε. But it has been demonstrated that if $\overline{\delta}_t^{(k)}(n_t) = 0$ $\forall n_t \in \mathcal{N}_t^{(k)}$ $(t = 2, ..., T)$, or equivalently, if $\delta_t^{(k)}(\xi_t) = 0$ (a.s.) $(t = 2, ..., T)$, then in general, the first stage candidate solution $\widehat{x}_1$ does not need to be optimal. Hence the information coming from the LP's (5.23) can be rather limited.*

Multiple refinements

It is certainly profitable to split simultaneously several nodes in `Step 2` of `MSLP-APPROX` in order to improve the efficiency. Otherwise, a subsequent master problem `MSLP-`$\mathcal{P}(\widehat{\mathcal{F}}^{(k+1)})$ can be too close to `MSLP-`$\mathcal{P}(\widehat{\mathcal{F}}^{(k)})$. On the other hand, of course, multiple refinements can be adverse to the increase of the node set $\mathcal{N}^{(k)}$ belonging to $\widehat{\mathcal{F}}^{(k)}$. And note then, in general, *the order* of node splitting also contributes to the shape of the next subfiltration $\widehat{\mathcal{F}}^{(k+1)}$. The latter problem does not arise if refinements take place at only one stage $\overline{t} = \overline{t}(k) \in \{2, ..., T\}$ per iteration k, whereas the associated rectangles $[\![\alpha_{\overline{t}}^{(n)}, \beta_{\overline{t}}^{(n)})\!)$, $n \in \mathcal{N}_{\overline{t}}^{(k)}$, are split *at most* in one coordinate $\overline{i} = \overline{i}(n)$. In the following alternative version of `Step 2`, some splittings are indicated at that stage $\overline{t} = \overline{t}(k)$ which takes on the highest number of assignments (5.19), and the affected nodes $\overline{n} \in \mathcal{N}_{\overline{t}}^{(k)}$ are split at the coordinate $\overline{i} = \overline{i}(\overline{n})$ according to the highest number of assignments. Starting from the unchanged `Step 0` and `Step 1` of `MSLP-APPROX` we reformulate

`Step 2:` (Multiple refinements)

- *Set* $J := \big\{1 \leq j \leq N_k \bigm| \Delta^{(k)}(\overline{\xi}_2^{(j)}, ..., \overline{\xi}_T^{(j)}) > \varepsilon\big\}$.
- $\forall j \in J$:
  $\Big[$ *Let* $(\overline{\xi}_2^{(j)}, ..., \overline{\xi}_T^{(j)})$ *be assigned to the node path* $(\overline{n}_2^{(j)}, ..., \overline{n}_T^{(j)})$ *associated with the sequence of expected vectors* $(\overline{\xi}_2^{[\overline{n}_2^{(j)}]}, ..., \overline{\xi}_T^{[\overline{n}_T^{(j)}]})$ *(cf. (5.12) and (5.9)). Select*

$$(\overline{t}^{(j)}, \overline{i}^{(j)}) \in \underset{2 \leq t \leq T,\ 1 \leq i \leq l_t}{\operatorname{argmax}} \left\{ \lambda_{ti} |\overline{\xi}_{ti}^{(j)} - \overline{\xi}_{ti}^{[\overline{n}_t^{(j)}]}| \right\} \quad \Big]. \tag{5.24}$$

- *Choose* $\bar{t} \in \mathrm{argmax}_{t\in\{2,...,T\}} \sharp\{j \in J \mid \bar{t}^{(j)} = t\}$.
  $\forall\, \overline{n} \in \mathcal{N}_{\bar{t}}^{(k)}$:
  $\Big[$ - *Set* $J_{\overline{n}} := \{j \in J \mid \overline{n}_{\bar{t}}^{(j)} = \overline{n}\}$.
  - *If* $J_{\overline{n}} \neq \emptyset$, *then select* $\bar{i} \in \mathrm{argmax}_{i\in\{1,...,l_{\bar{t}}\}} \sharp\{j \in J_{\overline{n}} \mid \bar{i}^{(j)} = i\}$ *and refine* $\widehat{\mathcal{F}}^{(k)}$ *according to the selection in (5.3) by*

$$t := \bar{t},\ n := \overline{n},\ i := \bar{i},\ \gamma := \frac{1}{2}(\alpha_{\bar{t}\,\bar{i}}^{[\overline{n}]} + \beta_{\bar{t}\,\bar{i}}^{[\overline{n}]}) \quad \Big].$$

- *Having the total refinement* $\widehat{\mathcal{F}}^{(k+1)}$, *set* $k := k + 1$ *and go to* `Step 1`.

Of course, the statement in Proposition 5.3 still remains valid by using this alternative `Step 2` because at least one rectangle is proposed to be split. But it is hardly possible to quantify a reduction of the iteration number. The difficulty comes mainly from the fact that the error function $\Delta^{(k)}$ does not need to be pointwise monotonically decreasing in k, as already mentioned earlier. When some simulated errors turn out to be small within a certain region of the random support, a similar behavior is basically not assured at the subsequent iteration.

5.3 Convergence to optimal solutions

The stopping criterion in `Step 1` of `MSLP-APPROX` is due to a scenario sample $(\overline{\xi}_2^{(j)}, ..., \overline{\xi}_T^{(j)})_{j=1,...,N_k}$ having that

$$\tfrac{1}{N_k} \textstyle\sum_{j=1}^{N_k} \Delta^{(k)}(\overline{\xi}_2^{(j)}, ..., \overline{\xi}_T^{(j)}) \leq \varepsilon \tag{5.25}$$

where the sample is independent of all previous samples at iterations $0 \leq k' < k$. Recall that the sample size at iteration k is given by $N_k := \lceil N e^{\varrho k} \rceil$ using the parameters $N \in \mathbb{N}$ and $\varrho > 0$. The stopping criterion can be interpreted as accepting the hypothesis

$$\mathcal{H}: \quad \mathbb{E}\Delta^{(k)}(\xi_2, ..., \xi_T) < \overline{\varepsilon} \tag{5.26}$$

where $\overline{\varepsilon} > \varepsilon > 0$. The significance level of $\mathcal{H}$ is connected with $\overline{\varepsilon}, \varepsilon, N, \varrho$ and k. The validity of (5.26) will be used to prove convergence to optimal solutions with respect to $\overline{\varepsilon} \downarrow 0$. Because `MSLP-APPROX` relies on simulation, the stopping time $\overline{k}$ is stochastic even though there exists a deterministic upper bound for $\overline{k}$ according to Proposition 5.3. When $\overline{k}$ is known, denoting

$$\mu_k := \mathbb{E}\Delta^{(k)}(\xi_2, ..., \xi_T) \ \ (k = 0, ..., \overline{k}),$$

the iterations of $K_0 := \{0 \le k \le \overline{k} \mid \mu_k \ge \overline{\varepsilon}\}$ must be regarded as critical since for $k \in K_0$, the algorithm took a chance to stop without having that $\mathbb{E}\Delta^{(k)}(\xi_2, ..., \xi_T) < \overline{\varepsilon}$. Let us define the aposteriori error by

$$\alpha_0 := \mathbb{P}\left[\min_{k\in K_0}\left\{\tfrac{1}{N_k}\textstyle\sum_{j=1}^{N_k}\Delta^{(k)}(\xi_2^{(j)}, ..., \xi_T^{(j)})\right\} \le \varepsilon\right], \tag{5.27}$$

where $\left\{(\xi_2^{(j)}, ..., \xi_T^{(j)})\right\}_{j=1,...,N_k,\, k\in K_0}$ are independent and identical distributed to $(\xi_2, ..., \xi_T)$. If α_0 is small, then one may reject the null hypothesis '$\overline{k} \in K_0$', or in other words, $\mathcal{H}$ is significant enough at the stopping time $k = \overline{k}$. The next lemma shows to what extent α_0 is small, independently of the stopping time $\overline{k}$.

Lemma 5.5
There is a model constant $\nu \in \mathbb{R}_+$ such that for the parameters $\varepsilon > 0, N \in \mathbb{N}\backslash\{0\}, \varrho > 0$ in `Step 0` of `MSLP-APPROX` and for $0 < \varepsilon < \overline{\varepsilon}$ it follows that

$$\alpha_0 < \frac{\nu}{N(1-e^{-\varrho})(\overline{\varepsilon}-\varepsilon)^2}. \tag{5.28}$$

Proof. By (5.16) we also have

$$\Delta^{(k)}(\xi_2, ..., \xi_T) \le \overline{C}\textstyle\sum_{t=2}^{T}\sum_{i=1}^{l_t}|\xi_{ti} - \widehat{\xi}_{ti}^{(k)}| \le \overline{C}\sum_{t=2}^{T}\sum_{i=1}^{l_t}(\beta_{ti} - \alpha_{ti}), \ \text{(a.s.)}, \tag{5.29}$$

this means that $\Delta^{(k)}$ is essentially bounded uniformly in k. It immediately follows that the variance is also uniformly bounded, i.e., there is a $\nu \in \mathbb{R}_+$ such that

$$\mathrm{var}\big(\Delta^{(k)}(\xi_2, ..., \xi_T)\big) \le \nu, \ \forall k.$$

Now let $k \in K_0$, i.e., $\mathbb{E}\Delta^{(k)}(\xi_2, ..., \xi_T) = \mu_k \ge \overline{\varepsilon}$. Chebyshev's inequality yields

$$\mathbb{P}\left[\left|\tfrac{1}{N_k}\textstyle\sum_{j=1}^{N_k}\Delta^{(k)}(\xi_2^{(j)}, ..., \xi_T^{(j)}) - \mu_k\right| \ge \delta\right] \le \frac{\nu}{N_k\delta^2}, \ \forall\delta > 0.$$

By choosing $\delta := \mu_k - \varepsilon \ge \overline{\varepsilon} - \varepsilon > 0$, we conclude that

$$\mathbb{P}\left[\tfrac{1}{N_k}\textstyle\sum_{j=1}^{N_k}\Delta^{(k)}(\xi_2^{(j)}, ..., \xi_T^{(j)}) \le \varepsilon\right] \le \frac{\nu}{N_k\delta^2} \le \frac{\nu}{N_k(\overline{\varepsilon}-\varepsilon)^2},$$

and it follows

$$\begin{aligned}
\alpha_0 &= \mathbb{P}\left[\min_{k\in K_0}\left\{\tfrac{1}{N_k}\textstyle\sum_{j=1}^{N_k}\Delta^{(k)}(\xi_2^{(j)},...,\xi_T^{(j)})\right\}\leq\varepsilon\right] \\
&\leq \textstyle\sum_{k\in K_0}\mathbb{P}\left[\tfrac{1}{N_k}\sum_{j=1}^{N_k}\Delta^{(k)}(\xi_2^{(j)},...,\xi_T^{(j)})\leq\varepsilon\right]\leq\sum_{k\in K_0}\dfrac{\nu}{N_k(\overline{\varepsilon}-\varepsilon)^2} \\
&= \textstyle\sum_{k\in K_0}\dfrac{\nu}{\lceil Ne^{\varrho k}\rceil(\overline{\varepsilon}-\varepsilon)^2}\leq\sum_{k\in K_0}\dfrac{\nu}{Ne^{\varrho k}(\overline{\varepsilon}-\varepsilon)^2} \\
&= \dfrac{\nu}{N(\overline{\varepsilon}-\varepsilon)^2}\textstyle\sum_{k\in K_0}\left(\frac{1}{e^{\varrho}}\right)^k<\dfrac{\nu}{N(\overline{\varepsilon}-\varepsilon)^2}\sum_{k=0}^{\infty}\left(\frac{1}{e^{\varrho}}\right)^k \\
&= \dfrac{\nu}{N(\overline{\varepsilon}-\varepsilon)^2(1-e^{-\varrho})}.
\end{aligned}$$

This completes the proof. □

An infinite version of `MSLP-APPROX` is now obtained by

<u>`MSLP-SOLVE`</u>

- *Choose the sequences* $(\varepsilon^{(r)}>0)_{r\geq1}$ *and* $(N^{(r)}\in\mathbb{N})_{r\geq1}$ *so that* $\varepsilon^{(r)}\overset{r\to\infty}{\longrightarrow}0$ *and* $\sum_{r=1}^{\infty}\frac{1}{N^{(r)}}<+\infty$.
- *For* $r=1,2,3,...$
  [– *apply* `MSLP-APPROX` *to the parameters* $\varepsilon:=\varepsilon^{(r)}$ *and* $N:=N^{(r)}$, *whereas the sample size increment* $\varrho>0$ *is assumed to be fixed;*
  – *let* $\widehat{\mathcal{F}}^{(r)}$ *be the subfiltration at the stopping time (which exists w.p.1 due to Prop. 5.3), associated with the complementarity variable* $\Delta^{(r)}(\xi_2,...,\xi_T)=\widehat{s}^{(r)\top}\overline{x}^{(r)}$ *where* $\overline{x}^{(r)}$ *is the feasible solution in* `MSLP-`$\mathcal{P}(\mathcal{F})$ *recursively defined by* $\widehat{x}^{(r)}\in\operatorname{argmin}\big($`MSLP-`$\mathcal{P}(\widehat{\mathcal{F}}^{(r)})\big)$ *and* $(\widehat{u}^{(r)},\widehat{s}^{(r)})\in\operatorname{argmax}\big($`MSLP-`$\mathcal{D}(\widehat{\mathcal{F}}^{(r)})\big)$.]

Proposition 5.6
The outcomes of `MSLP-SOLVE` *imply (a)-(c) w.p.1:*

(a) *It holds* $\lim_{r\to\infty}\mathbb{E}\Delta^{(r)}(\xi_2,...,\xi_T)=\lim_{r\to\infty}\mathbb{E}[\widehat{s}^{(r)\top}\overline{x}^{(r)}]=0$ *and, therefore,*

$$\lim_{r\to\infty}\mathbb{E}[\mathbf{b}^{\top}\widehat{u}^{(r)}]=\sup\big(\texttt{MSLP-}\mathcal{D}(\mathcal{F})\big)=\inf\big(\texttt{MSLP-}\mathcal{P}(\mathcal{F})\big)=\lim_{r\to\infty}\mathbb{E}[\mathbf{c}^{\top}\overline{x}^{(r)}].$$

(b) *Every weak accumulation point (in* L^2*) of* $\{\overline{x}^{(r)}\}_{r\geq1}$ *is an optimal solution to* `MSLP-`$\mathcal{P}(\mathcal{F})$. *Such points exist whenever (5.1) is satisfied.*

(c) *Every weak accumulation point (in L^2) of $\{(\widehat{u}^{(r)}, \widehat{s}^{(r)})\}_{r\geq 1}$ is an optimal solution to* `MSLP-`$\mathcal{D}(\mathcal{F})$*, and such points always exist.*

Proof. (a): Because `MSLP-APPROX` relies on simulation, the chain $\{\widehat{\mathcal{F}}^{(r)}\}_{r\geq 1}$ is realized in an infinite sample space and thus also $\{\Delta^{(r)}\}_{r\geq 1}$. Let $\overline{\varepsilon} > 0$ be arbitrary having that $\varepsilon^{(r)} \leq \frac{\overline{\varepsilon}}{2}$ for all $r \geq \overline{r}$. If $\mathbb{E}\Delta^{(r)}(\xi_2, ..., \xi_T) \geq \overline{\varepsilon}$ for $r \geq \overline{r}$, then the subroutine `MSLP-APPROX` of pass r would mistakenly accept the hypothesis $\mathcal{H}$ in (5.26). This happens due to Lemma 5.5 with probability $\alpha_0^{(r)}$ estimated from above by

$$\alpha_0^{(r)} < \frac{\nu}{N^{(r)}(1-e^{-\varrho})(\overline{\varepsilon}-\varepsilon^{(r)})^2} \leq \frac{4\nu}{N^{(r)}(1-e^{-\varrho})\overline{\varepsilon}^2}.$$

Since $\sum_{r=1}^{\infty} \frac{1}{N^{(r)}} < +\infty$ by assumption, we conclude that $\sum_{r=1}^{\infty} \alpha_0^{(r)} < +\infty$, and hence, Borel-Cantelli's Lemma implies that, w.p.1, $\mathbb{E}\Delta^{(r)}(\xi_2, ..., \xi_T) \leq \overline{\varepsilon}$ for all r large enough. Since $\overline{\varepsilon} > 0$ was arbitrary, it follows that $\lim_{r\to\infty} \mathbb{E}\Delta^{(r)}(\xi_2, ..., \xi_T) = 0$. For the second part of (a) we refer back to (2.11).

(b) & (c): Referring to the short form of `MSLP-`$\mathcal{P}(\mathcal{F})$ and `MSLP-`$\mathcal{D}(\mathcal{F})$ in (2.7) and (2.8), respectively, we note that the linear mappings $\mathbf{A}$ and $\mathbf{A}^*$ are continuous. Hence the feasible sets of the problems are closed (in L^2). Since they are also convex, Mazur's Theorem (see e.g. [12]) states that they are also weakly closed. Let $\overline{\overline{x}}$ be a weak accumulation point of $\{\overline{x}^{(r)}\}_{r\geq 1}$, that is therefore feasible in `MSLP-`$\mathcal{P}(\mathcal{F})$ and there is a weakly convergent subsequence $\{\overline{x}^{(r_k)}\}_{k\geq 1}$ with $\overline{x}^{(r_k)} \rightharpoonup \overline{\overline{x}}$, i.e., $\lim_{k\to\infty} \mathbb{E}[z^\top \overline{x}^{(r_k)}] = \mathbb{E}[z^\top \overline{\overline{x}}]$, $\forall z \in L^2(\mathcal{F}; \mathbb{R}^n)$. In particular, it holds $\mathbb{E}[\mathbf{c}^\top \overline{\overline{x}}] = \lim_{k\to\infty} \mathbb{E}[\mathbf{c}^\top \overline{x}^{(r_k)}] \overset{\text{(a)}}{=} \inf(\texttt{MSLP-}\mathcal{P}(\mathcal{F}))$ and thus, $\overline{\overline{x}}$ is even an optimal solution to `MSLP-`$\mathcal{P}(\mathcal{F})$. A similar argument applies to (c), this means that every weak accumulation point of $\{(\widehat{u}^{(r)}, \widehat{s}^{(r)})\}_{r\geq 1}$ solves `MSLP-`$\mathcal{D}(\mathcal{F})$.

Furthermore, assumption (5.1) allows to apply Lemma 4.8, therefore $\{\overline{x}^{(r)}\}_{r\geq 1}$ is bounded in L^∞ and so in L^2. Hence there exists at least one weak accumulation point of $\{\overline{x}^{(r)}\}_{r\geq 1}$ (see [12], p.187). A similar argument applies for $\{(\widehat{u}^{(r)}, \widehat{s}^{(r)})\}_{r\geq 1}$ using instead Lemma 4.11. ☐

6 Numerical results

The primary goal of our numerical tests is to get an impression of the practical behavior of the algorithm `MSLP-APPROX` proposed in Section 5.2. In particular, the focus is on the improvement of the bounds due to refinements. We first discuss some general facets of the implementation. Section 6.2 presents numerical results for the inventory model introduced in Section 1.1 and extended to T stages. In Section 6.3, `MSLP-APPROX` is tested for a randomly generated multistage problem by varying some model parameters.

6.1 General issues concerning the implementation

The method `MSLP-APPROX` has been implemented and tested in the general algebraic modeling language `GAMS`, see Brooke et al. [6]. The implementation is available for (truncated) normal distributed data. All the linear master- and subproblems (5.10)-(5.11) and (5.13) have been solved with `GAMS/CPLEX`, whereas the solver `GAMS/MINOS` has been used for the quadratic subproblems (5.14). The refinement procedure in `Step 2` of `MSLP-APPROX` has been implemented according to the scheme on page 64.

Because the infinite extension `MSLP-SOLVE` of Section 5.3 refers solely to a proof of convergence to optimal solutions, we content ourselves with `MSLP-APPROX` by using a fixed splitting tolerance $\varepsilon > 0$. In view of Proposition 5.3, the reason is that the computational effort can increase exponentially if ε tends to 0. In particular, we choose

$$\varepsilon := \theta \cdot (|\mathrm{LB}^{(0)}| + 1)$$

where e.g. $\theta = 0.01$ and $\mathrm{LB}^{(0)} := \min\big(\texttt{MSLP-}\mathcal{P}(\mathcal{T})\big)$ denotes the optimal value of the expected value problem, i.e., $\mathrm{LB}^{(0)}$ is the lower bound resulting from iteration $k = 0$. In order to compare the improvement of the bounds, the focus here is on the estimated expected complementarity variable

$$\overline{\Delta}^{(k)} := \tfrac{1}{N_k} \textstyle\sum_{j=1}^{N_k} \Delta^{(k)}(\overline{\xi}_2^{(j)}, ..., \overline{\xi}_T^{(j)})$$

at iterations $k = 0, 1, 2, 3...$, rather than at the stopping time $\overline{k}$. In particular, the stopping criterion $\overline{\Delta}^{(k)} \leq \varepsilon$ is ignored here as long as there is a j with

$\Delta^{(k)}(\overline{\xi}_2^{(j)}, ..., \overline{\xi}_T^{(j)}) > \varepsilon$. The estimated standard error of the mean $\overline{\Delta}^{(k)}$ is specified as

$$\mathrm{SE}^{(k)} := \frac{1}{\sqrt{N_k}} \sqrt{\frac{\sum_{j=1}^{N_k} \left(\Delta^{(k)}(\overline{\xi}_2^{(j)}, ..., \overline{\xi}_T^{(j)}) - \overline{\Delta}^{(k)} \right)^2}{N_k - 1}}. \tag{6.1}$$

Assuming the setting of page 53 with regard to (5.29), the variable $\Delta^{(k)}(\xi_2, ..., \xi_T)$ is essentially bounded uniformly in k, i.e., there is a $\mathcal{C} \in \mathbb{R}_+$ such that $0 \leq \Delta^{(k)}(\xi_2, ..., \xi_T) \leq \mathcal{C}$ (a.s.) for all k. Hence the variance is bounded by

$$\operatorname{var}\left[\Delta^{(k)}(\xi_2, ..., \xi_T)\right] \leq \mathbb{E}\left[\Delta^{(k)}(\xi_2, ..., \xi_T)^2\right] \leq \mathcal{C}\mathbb{E}\Delta^{(k)}(\xi_2, ..., \xi_T)$$

and therefore is at least linearly decreasing with the decrease of the gap $\mathbb{E}\Delta^{(k)}(\xi_2, ..., \xi_T)$. Thus, a reduction of the standard error (6.1) results from both an increase of the sample size and a decrease of the expected complementarity variable.

It should be noted that our implementation does not utilize the following two facilities which can significantly improve the efficiency:

- The computation of one single value $\Delta^{(k)}(\overline{\xi}_2^{(j)}, ..., \overline{\xi}_T^{(j)})$ according to (5.15) involves recursively the evaluation of $(T-1)$ LP‘s together with $(T-2)$ QP‘s. However, the loop of size N_k for evaluating the mean $\overline{\Delta}^{(k)}$ can be parallelized because different scenarios feature independent subproblems.

- Each iteration k begins with solving the primal-dual master problems $\texttt{MSLP-}\mathcal{P}(\widehat{\mathcal{F}}^{(k)})$ and $\texttt{MSLP-}\mathcal{D}(\widehat{\mathcal{F}}^{(k)})$. Depending on the LP-solver, the solution of the one problem gives simultaneously the solution of the other. Moreover, an optimal dual solution $(\widehat{u}, \widehat{s}) \in \operatorname{argmax}\big(\texttt{MSLP-}\mathcal{D}(\widehat{\mathcal{F}}^{(k)})\big)$ could be used as a feasible starting solution in the next problem $\texttt{MSLP-}\mathcal{D}(\widehat{\mathcal{F}}^{(k+1)})$ (cf. Theorem 3.14).

6.2 A T-stage inventory model

We are dealing with the inventory model introduced in Section 1.1 and extended here to T stages. For simplicity, we suppose that the product under consideration has a linearly increasing market price

$$p_t := 1 + \tfrac{t-1}{T-1}\delta \quad (t = 1, ..., T) \tag{6.2}$$

where $\delta > 0$ is the total markup from $t = 1$ to $t = T$. Furthermore, the demand on the product at time $t = 2, ..., T$ is assumed to be given by the martingale $(\eta_t)_{t=2,...,T}$,

$$\eta_t := \eta_{t-1} + \xi_t = \eta_1 + \textstyle\sum_{s=2}^{t} \xi_s, \tag{6.3}$$

where $\xi_2, ..., \xi_T \sim \frac{\sigma}{\sqrt{T-1}} N(0,1)$ are i.i.d. and $\eta_1 := 100$. Note that the variance $\mathrm{var}(\eta_T) = \mathrm{var}\big(100 + \sum_{t=2}^{T} \xi_t\big) = \sigma^2$ of the last demand is independent of T. For this reason, (6.2) and (6.3) correspond rather to a time discretized model than to a consideration of what happens if the time horizon goes to infinity; the latter is anyway of minor interest in the area of middle-term planning. We now extend the three-stage model (1.5) to

$$\begin{aligned} \mathcal{Q}_1 = \min_{\mathtt{x}_1, \mathtt{y}_1} \;\; & p_1 \mathtt{x}_1 + \mathbb{E}[\mathcal{Q}_2(\mathtt{y}_1, \eta_2)] \\ \text{s.t. } \; & \mathtt{y}_1 - \mathtt{x}_1 = 0 \\ & \mathtt{y}_1 \geq 0 \end{aligned} \tag{6.4}$$

where for $t = 2, ..., T$

$$\begin{aligned} \mathcal{Q}_t\big(\mathtt{y}_{t-1}; \eta_t\big) := \min_{\mathtt{x}_t, \mathtt{y}_t} \;\; & p_t \mathtt{x}_t + \mathbb{E}\big[\mathcal{Q}_{t+1}(\mathtt{y}_t, \eta_{t+1}) \bigm| \eta_t\big] \\ \text{s.t. } \; & \mathtt{y}_t - \mathtt{x}_t = \mathtt{y}_{t-1} \\ & \mathtt{x}_t \geq -\eta_t \\ & \mathtt{y}_t \geq 0 \end{aligned}$$

with $\mathcal{Q}_{T+1} \equiv 0$. Recall that the goal in (6.4) is to minimize the negative expected profit where the feasible activities are summarized as follows:

- $\mathtt{x}_t \geq 0$ means buying and $\mathtt{x}_t \leq 0$ means selling the amount of $|\mathtt{x}_t|$ at time $t = 1, ..., T$, in both cases at price p_t; note that $\mathtt{x}_1 \geq 0$ is enforced, i.e., the procedure begins with buying at price $p_1 = 1$;

- $\mathtt{y}_t = \mathtt{y}_{t-1} + \mathtt{x}_t$ denotes the stock at the end of time $t = 2, ..., T$, where $\mathtt{y}_1 = \mathtt{x}_1$;

- the restrictions $\mathtt{y}_t \geq 0$ and $\mathtt{x}_t \geq -\eta_t$ are solely for physical reasons; the latter prevents to sell more than the demand permits at time t.

The infinite LP formulation of (6.4) turns out to be

$$\texttt{MSLP-}\mathcal{P}: \qquad \text{Minimize}_{x,y} \ \mathbb{E}\left[\sum_{t=1}^{T} p_t x_t\right]$$

$$\text{subject to} \left\{ \begin{array}{rcll} y_1 - x_1 & = & 0 & \\ -y_{t-1} + y_t - x_t & = & 0 \ \text{(a.s.)} & (t = 2, ..., T) \\ x_t & \geq & -\eta_t \ \text{(a.s.)} & (t = 2, ..., T) \\ y_t & \geq & 0 \ \text{(a.s.)} & (t = 1, ..., T), \end{array} \right. \tag{6.5}$$

where $x_1, y_1 \in \mathbb{R}$ and $x_t = x_t(\eta_2, ..., \eta_t),\ y_t = y_t(\eta_2, ..., \eta_t),\ (t = 2, ..., T)$. Note that the integrand of the total negative profit can also be written as

$$\begin{aligned} \textstyle\sum_{t=1}^{T} p_t x_t &= p_1 y_1 + \textstyle\sum_{t=2}^{T} p_t \big(y_t - y_{t-1}\big) = \textstyle\sum_{t=1}^{T-1} (p_t - p_{t+1}) y_t + p_T y_T \\ &\overset{(6.2)}{=} \textstyle\sum_{t=1}^{T-1} \left(\frac{-\delta}{T-1}\right) y_t + (1 + \delta) y_T. \end{aligned}$$

We reformulate '$x_t \geq -\eta_t$' as '$-y_{t-1} + y_t \geq -\eta_t$', leading equivalently to '$-y_{t-1} + y_t - z_t = -\eta_t,\ z_t \geq 0$' by adding a slack variable $z_t = z_t(\eta_2, ..., \eta_t)$. Furthermore, with regard to (6.3), the demand process $(\eta_2, ..., \eta_T)$ is determined by the noise process $(\xi_2, ..., \xi_T)$. Formulation (6.5) is therefore equivalently rewritten in standard form as

$$\texttt{MSLP-}\mathcal{P}: \qquad \text{Minimize}_{y,z} \ \mathbb{E}\left[\sum_{t=1}^{T-1} \left(\frac{-\delta}{T-1}\right) y_t + (1 + \delta) y_T\right]$$

$$\text{subject to} \left\{ \begin{array}{rcll} -y_{t-1} + y_t - z_t & = & -\big(100 + \sum_{s=2}^{t} \xi_s\big) \ \text{(a.s.)} & (t = 2, ..., T) \\ y_t, z_t & \geq & 0 \ \text{(a.s.)} & (t = 1, ..., T)\ , \end{array} \right. \tag{6.6}$$

where $y_1 \in \mathbb{R}$ and $y_t = y_t(\xi_2, ..., \xi_t),\ z_t = z_t(\xi_2, ..., \xi_t),\ (t = 2, ..., T)$. There are no equality constraints at stage 1, whereas in the terminology of (2.6) one has

$$A_t = (1, -1),\ B_t = (-1, 0),\ b_t = -\big(100 + \textstyle\sum_{s=2}^{t} \xi_s\big),\ (t = 2, ..., T),$$

$c_t^\top = (\frac{-\delta}{T-1}, 0)$ $(t = 1, ..., T-1)$ and $c_T^\top = (1 + \delta, 0)$. Because the method `MSLP-APPROX` is restricted to random variables with bounded support, we use the truncated normal distribution $\xi_t \sim \frac{\sigma}{\sqrt{T-1}} N(0,1)_{[-4,4]}$ where $N(0,1)_{[-4,4]}$ denotes the standard normal distribution conditioned on the interval $[-4, 4]$. The only parameters occurring in (6.6) are $\sigma > 0$, $\delta > 0$ and $T \in \{2, 3, ...\}$. All assumptions M1)-M3) of page 53 are satisfied for (6.6); in particular, the problem is RCR^o because $A_t \mathbb{R}_+^2 = \mathbb{R}$ $(t = 2, ..., T)$, whereas (5.1) is superfluous since only $(b_t)_{t=2,...,T}$ is random.

The computational results of `MSLP-APPROX` for $T = 2, 3, 5, 10$ and 20 are shown in Table 6.1 for $\sigma = 10$ and Table 6.2 using $\sigma = 30$. The markup in (6.2) has always been taken as $\delta = 0.5$. The parameters in `Step 0` of `MSLP-APPROX` have been chosen as

$$\varepsilon := 0.01 \cdot (|\mathrm{LB}^{(0)}| + 1),\ N := 100,\ \varrho := 0.05,\ \lambda_{t1} = 1\ \forall t,$$

where $\mathrm{LB}^{(0)} = \min\big(\texttt{MSLP-}\mathcal{P}(\mathcal{T})\big)$ is the optimal value of the expected value problem. The column headers in Table 6.1 and 6.2 have the following meaning:

- k: iteration;
- $\sharp\mathcal{N}_T^{(k)}$: number of leaf nodes (approximating scenarios);
- $N_k = \lceil N e^{\varrho k} \rceil$: sample size;
- $\widehat{y}_1^{(k)}$: optimal first stage solution of $\texttt{MSLP-}\mathcal{P}(\widehat{\mathcal{F}}^{(k)})$ (quantity to buy at stage 1);
- $\mathrm{LB}^{(k)} := \min\big(\texttt{MSLP-}\mathcal{P}(\widehat{\mathcal{F}}^{(k)})\big)$: lower bound;
- $\overline{\Delta}^{(k)} := \frac{1}{N_k}\sum_{j=1}^{N_k} \Delta^{(k)}(\overline{\xi}_2^{(j)}, ..., \overline{\xi}_T^{(j)})$: unbiased estimator of the duality gap;
- $\overline{\mathrm{UB}}^{(k)} := \mathrm{LB}^{(k)} + \overline{\Delta}^{(k)}$: estimated upper bound (see also Section 3.5);
- $2 * \mathrm{SE}^{(k)}$: two times the standard error of the mean $\overline{\Delta}^{(k)}$ according to (6.1);
- $\frac{\overline{\Delta}^{(k)}}{|\mathrm{LB}^{(k)}|+1} = \frac{\overline{\mathrm{UB}}^{(k)} - \mathrm{LB}^{(k)}}{|\mathrm{LB}^{(k)}|+1}$: relative error of the bounds;
- $\bar{t}$: splitting stage (see page 64).

We make the following comments on the results:

- Concerning the expected value problems (iteration $k = 0$), the optimal quantity $\widehat{y}_1^{(k)}$ to buy at the first stage is $100(T-1)$. This corresponds to the selling of the constant amount of 100 per stage $t = 2, ..., T$ without having a surplus at the end with respect to the expected demands. It is worth mentioning that, in general, the sequence of the candidate solutions $(\widehat{y}_1^{(k)})_{k\geq 0}$ resulting from refinement is not monotone.

- The instance whith $T = 2$ corresponds to the newsvendor problem of Section 1.1. Convergence is achieved very fast with only one splitting per iteration. In both cases, $\sigma = 10$ and $\sigma = 30$, the support $[-4, 4]$ of $\xi_2 \sim N(0,1)_{[-4,4]}$ has been split in the order

$$\begin{aligned} [-4,4] &\xrightarrow{k=1} \big([-4,0],[0,4]\big) \xrightarrow{k=2} \big([-4,-2],[-2,0],[0,4]\big) \\ &\xrightarrow{k=3} \big([-4,-2],[-2,-1],[-1,0],[0,4]\big) \\ &\xrightarrow{k=4} \big([-4,-2],[-2,-1],[-1,-\tfrac{1}{2}],[-\tfrac{1}{2},0],[0,4]\big), \text{ etc.} \end{aligned}$$

- In the instances where $T = 2, 3$ and 5, the relative error $\frac{\overline{\Delta}^{(k)}}{|\mathrm{LB}^{(k)}|+1}$ of the bounds decreases within a moderate effort to a quite small value (see columns next to last). Having more stages, the fast growing size of the master problems makes further significant improvements computationally very expensive. For the instance where $T = 20$ and $\sigma = 30$ (see Table 6.2), the relative error decreases from 0.386 only to 0.275 with more than 10^3 approximating scenarios. An improvement to a value of less than 0.100 would probably need an exorbitant number of nodes. Some frequency diagrams of the simulated values of $\Delta^{(k)}(\xi_2, ..., \xi_T)$ for $T = 5, 10$ and 20 are shown in Fig. 6.1. In the case $T = 5$, the number of simulated values $\Delta^{(25)}(\overline{\xi}_2^{(j)}, ..., \overline{\xi}_T^{(j)}) \approx 0$ is quite large (279/350 for $\sigma = 10$ and 281/350 for $\sigma = 30$). On the other hand, it is not surprising that the outliers are more numerous in the case $\sigma = 30$.

- For all the examples one observes that $\mathrm{LB}^{(k)} - \mathrm{LB}^{(0)} > \overline{\mathrm{UB}}^{(0)} - \overline{\mathrm{UB}}^{(k)}$ holds; this means that the improvement of the lower bound is greater than of the upper bound. Hence one might suppose that, in many cases, the upper bounds are closer to the unknown optimal objective value than the lower bounds. An illustration of the approximating scenario tree for $T = 3$ and $\sigma = 30$ at iteration 12 is given in Fig. 6.2, whereas Fig. 6.3 refers to the tree for $T = 5$ and $\sigma = 30$ at iteration 9. Note that the subtrees at some stage $2 \le t \le T - 1$ may have a different structure.

Table 6.1: Results for the inventory model (6.6) with $\delta = 0.5$ and $\sigma = 10$.

$\delta = 0.5$, $\sigma = 10$										
T	k	$\sharp\mathcal{N}_T^{(k)}$	N_k	$\widehat{y}_1^{(k)}$	$\text{LB}^{(k)}$	$\overline{\text{UB}}^{(k)}$	$\overline{\Delta}^{(k)}$	$\pm 2{*}\text{SE}^{(k)}$	$\frac{\overline{\Delta}^{(k)}}{\lvert\text{LB}^{(k)}\rvert+1}$	$\bar{t}$
2	0	1	100	100.000	-50.000	-43.339	6.661	0.944	0.131	2
	1	2	106	92.023	-46.012	-44.289	1.723	0.481	0.037	2
	2	3	111	92.772	-45.824	-44.233	1.592	0.410	0.034	2
	3	4	117	95.401	-45.167	-44.537	0.630	0.205	0.014	2
	4	5	123	97.552	-44.630	-44.420	0.210	0.094	0.005	2
	5	6	129	96.269	-44.582	-44.541	0.041	0.030	0.001	2
	6	**7**	**135**	**95.631**	**-44.558**	**-44.548**	**0.010**	**0.009**	**0.000**	**-**
3	0	1	100	200.000	-75.000	-64.696	10.304	1.582	0.136	2
	1	2	106	188.719	-72.180	-65.850	6.330	1.055	0.086	3
	2	4	111	194.360	-69.360	-67.812	1.548	0.658	0.022	3
	3	7	117	193.830	-69.227	-67.412	1.815	0.747	0.026	2
	4	14	123	184.138	-69.022	-66.602	2.420	0.788	0.035	2
	5	21	129	187.856	-68.735	-67.440	1.294	0.427	0.019	3
	10	84	165	190.436	-67.674	-67.541	0.133	0.090	0.002	3
	15	**134**	**212**	**189.313**	**-67.615**	**-67.562**	**0.053**	**0.038**	**0.001**	**-**
5	0	1	100	400.000	-125.000	-106.444	18.556	3.103	0.147	4
	1	2	106	392.023	-122.009	-110.876	11.132	2.083	0.090	2
	2	4	111	376.070	-120.015	-110.398	9.617	1.981	0.079	3
	3	8	117	380.058	-117.771	-110.291	7.480	1.390	0.063	5
	4	16	123	376.070	-116.649	-110.744	5.905	1.469	0.050	2
	5	32	129	377.568	-116.334	-112.394	3.940	1.099	0.034	5
	10	188	165	374.284	-115.148	-113.068	2.080	0.852	0.018	4
	15	474	212	377.007	-114.295	-113.257	1.038	0.488	0.009	5
	20	1379	272	375.403	-114.080	-113.285	0.795	0.303	0.007	4
	25	**1957**	**350**	**375.516**	**-113.994**	**-113.309**	**0.685**	**0.216**	**0.006**	**-**
10	0	1	100	900.000	-250.000	-214.150	35.850	6.203	0.143	3
	1	2	106	878.729	-247.637	-220.770	26.867	4.699	0.108	2
	2	4	111	854.799	-246.307	-220.367	25.940	4.751	0.105	10
	3	8	117	852.140	-244.978	-222.677	22.301	6.306	0.091	9
	4	14	123	852.140	-243.630	-225.271	18.359	5.452	0.075	6
	5	28	129	844.163	-240.574	-221.839	18.735	5.160	0.078	4
	10	372	165	849.481	-234.387	-225.934	8.453	2.248	0.036	8
	15	**1476**	**212**	**846.410**	**-233.694**	**-223.406**	**10.289**	**2.291**	**0.044**	**-**
20	0	1	100	1900.000	-500.000	-434.363	65.637	10.838	0.131	17
	1	2	106	1892.680	-496.918	-436.573	60.345	14.724	0.121	7
	2	4	111	1867.060	-492.873	-429.221	63.652	14.784	0.129	3
	3	8	117	1834.121	-491.139	-433.588	57.551	12.669	0.117	13
	4	14	123	1834.121	-488.876	-428.313	60.563	12.951	0.124	12
	5	28	129	1832.291	-485.673	-441.922	43.751	9.910	0.090	5
	10	382	165	1812.161	-478.399	-442.089	36.310	9.050	0.076	11
	12	**642**	**183**	**1812.161**	**-477.741**	**-443.421**	**34.320**	**7.316**	**0.072**	-

Table 6.2: Results for the inventory model (6.6) with $\delta = 0.5$ and $\sigma = 30$.

$\delta = 0.5$, $\sigma = 30$										
T	k	$\sharp\mathcal{N}_T^{(k)}$	N_k	$\widehat{y}_1^{(k)}$	$\text{LB}^{(k)}$	$\overline{\text{UB}}^{(k)}$	$\overline{\Delta}^{(k)}$	$\pm 2{*}\text{SE}^{(k)}$	$\frac{\overline{\Delta}^{(k)}}{\lvert\text{LB}^{(k)}\rvert+1}$	$\bar{t}$
2	0	1	100	100.000	-50.000	-30.017	19.983	2.831	0.392	2
	1	2	106	76.070	-38.035	-32.866	5.169	1.442	0.132	2
	2	3	111	78.316	-37.473	-32.698	4.776	1.230	0.124	2
	3	4	117	86.204	-35.501	-33.611	1.891	0.616	0.052	2
	4	5	123	92.655	-33.889	-33.259	0.630	0.281	0.018	2
	5	6	129	88.808	-33.746	-33.623	0.122	0.089	0.004	2
	6	7	135	86.892	-33.674	-33.645	0.029	0.026	0.001	2
	7	**8**	**142**	**87.816**	**-33.646**	**-33.640**	**0.006**	**0.010**	**0.000**	**-**
3	0	1	100	200.000	-75.000	-44.089	30.911	4.746	0.407	2
	1	2	106	166.158	-66.539	-47.550	18.990	3.166	0.281	3
	2	4	111	183.079	-58.079	-53.436	4.643	1.974	0.079	3
	3	8	117	181.491	-57.682	-52.236	5.446	2.240	0.093	2
	4	16	123	154.002	-57.040	-49.989	7.051	2.286	0.121	2
	5	24	129	165.157	-56.022	-52.277	3.745	1.245	0.066	3
	10	69	165	171.147	-52.932	-52.668	0.263	0.242	0.005	2
	15	162	212	168.100	-52.830	-52.705	0.125	0.087	0.002	2
	20	**268**	**272**	**168.232**	**-52.800**	**-52.734**	**0.066**	**0.045**	**0.001**	**-**
5	0	1	100	400.000	-125.000	-69.461	55.539	9.291	0.441	4
	1	2	106	376.070	-116.026	-82.684	33.342	6.218	0.285	2
	2	4	111	328.210	-110.044	-81.194	28.850	5.944	0.260	3
	3	8	117	340.175	-103.313	-80.874	22.440	4.170	0.215	5
	4	16	123	328.210	-99.948	-82.232	17.716	4.408	0.175	2
	5	32	129	332.703	-99.002	-87.181	11.821	3.297	0.118	5
	10	197	165	322.851	-95.354	-89.106	6.248	2.555	0.065	4
	15	848	212	326.210	-92.862	-89.621	3.240	1.508	0.035	4
	20	1695	272	325.470	-92.137	-90.085	2.052	0.840	0.022	4
	25	**1847**	**350**	**326.543**	**-92.018**	**-89.973**	**2.044**	**0.688**	**0.022**	**-**
10	0	1	100	900.000	-250.000	-144.149	105.851	18.429	0.422	7
	1	2	106	868.093	-239.364	-152.604	86.760	16.313	0.361	2
	2	4	111	796.303	-235.376	-154.878	80.499	15.552	0.341	10
	3	8	117	804.280	-232.163	-166.869	65.295	17.808	0.280	8
	4	16	123	812.257	-226.402	-160.849	65.554	16.669	0.288	6
	5	32	129	788.327	-220.531	-166.136	54.394	13.035	0.246	3
	10	374	165	732.490	-202.449	-172.910	29.539	7.660	0.145	9
	15	754	212	743.950	-201.170	-163.502	37.668	7.620	0.186	9
	20	**3638**	**272**	**736.982**	**-198.949**	**-168.601**	**30.348**	**6.256**	**0.152**	**-**
20	0	1	100	1900.000	-500.000	-306.417	193.583	32.267	0.386	17
	1	2	106	1878.040	-490.754	-311.183	179.570	42.561	0.365	20
	2	4	111	1883.530	-488.948	-285.277	203.671	49.452	0.416	7
	3	8	117	1806.671	-477.210	-296.799	180.410	41.180	0.377	13
	4	16	123	1806.671	-467.042	-299.762	167.280	39.169	0.357	4
	5	32	129	1713.343	-460.099	-330.833	129.266	34.635	0.280	5
	10	374	165	1614.524	-443.409	-332.250	111.159	25.567	0.250	11
	12	**1044**	**183**	**1598.054**	**-435.416**	**-315.207**	**120.209**	**25.651**	**0.275**	**-**

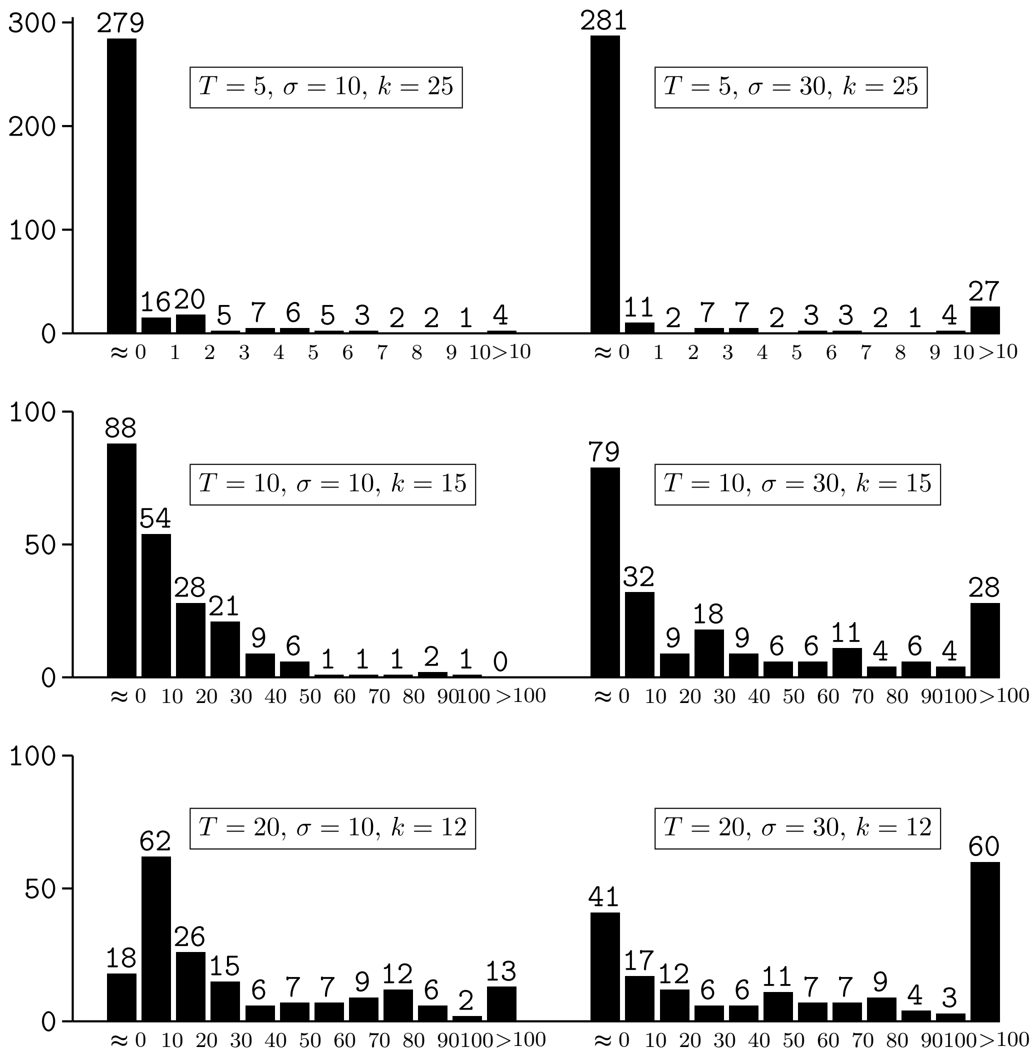

Figure 6.1: *Frequency of the simulated values of* $\Delta^{(k)}(\xi_2, ..., \xi_T)$.

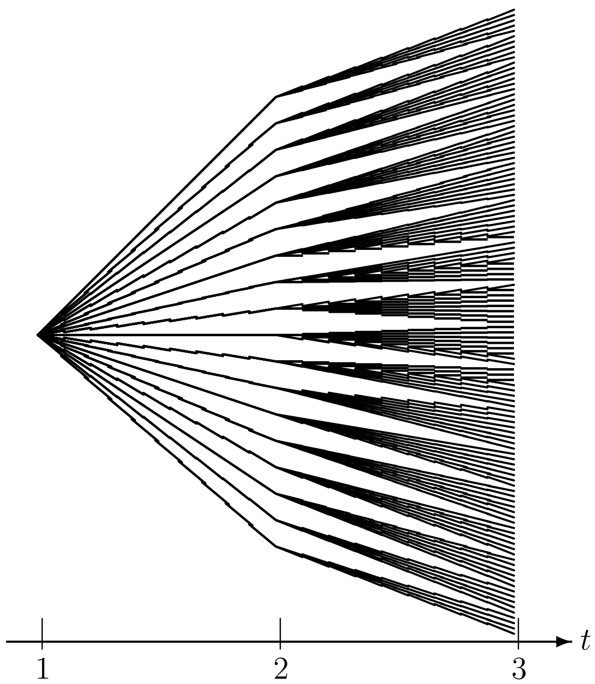

Figure 6.2: *Scenario tree for $T = 3$, $\sigma = 30$ and $k = 12$.*

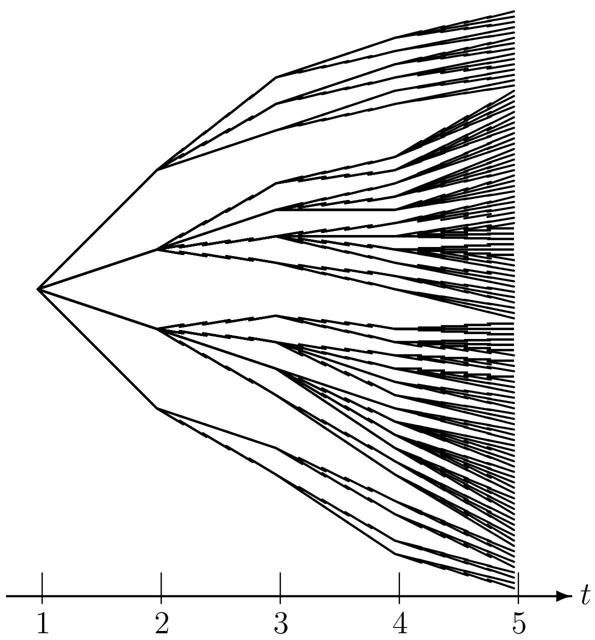

Figure 6.3: *Scenario tree for $T = 5$, $\sigma = 30$ and $k = 9$.*

6.3 A randomly generated problem

For generating a random recourse problem we have used the program `GenSLP` that is implemented in the model managment system `SLP-IOR` of Kall/Mayer [15]. `GenSLP` is a workbench for generating random vectors and matrices where both the density of nonzeros and the magnitude of the entries is adjustable. Optionally, the matrices $A \in \mathbb{R}^{m\times n}$ are built such that the complete recourse property $A\mathbb{R}^n_+ = \mathbb{R}^m$ holds. An algorithmic description can be found in Mayer [20]. The experiment here works as follows: using `GenSLP`, we have generated nonnegative cost vectors $(c_t \in \mathbb{R}^8)_{t=1,\ldots,5}$ and expected right-hand side vectors $(\mathtt{b}_t \in \mathbb{R}^4)_{t=1,\ldots,5}$ in addition to complete recourse matrices $(A_t \in \mathbb{R}^{4\times 8})_{t=1,..,5}$ and expected technology matrices $(\mathtt{B}_t \in \mathbb{R}^{4\times 8})_{t=2,\ldots,5}$ of density 60%. The generated data are shown in Fig. 6.4. Given 24 independent (truncated) standard normal variables $(\xi_{ti})_{i=1,..,6,\ t=2,\ldots,5}$, the stochastic parts have been modeled by

$$b_t := \mathtt{b}_t + \sigma_1 \begin{pmatrix} \xi_{t1} \\ \xi_{t2} \\ \xi_{t3} \\ \xi_{t4} \end{pmatrix} \quad (t = 2, ..., 5)$$

and

$$B_t := \mathtt{B}_t + \sigma_2 \left(\begin{array}{cccc|cccc} \xi_{t5} & \xi_{t5} & \xi_{t5} & \xi_{t5} & \xi_{t6} & \xi_{t6} & \xi_{t6} & \xi_{t6} \\ \vdots & & & \vdots & \vdots & & & \vdots \\ \xi_{t5} & \xi_{t5} & \xi_{t5} & \xi_{t5} & \xi_{t6} & \xi_{t6} & \xi_{t6} & \xi_{t6} \end{array}\right) \quad (t = 2, ..., 5)$$

where $\sigma_1, \sigma_2 \in \mathbb{R}$ are scale parameters. The computational results of `MSLP-APPROX` by using different numbers of stages $T = 2, ..., 5$ and scale parameters are shown in Table 6.3. In this respect, the problem data $(A_t, B_t, b_t, c_t)_{t=T+1,..,5}$ have been cut in the examples with $T < 5$. We compare the results for $\sigma_1 = 0.5$ and $\sigma_1 = 2$ where in each case either $\sigma_2 = 0$ (only b_t stochastic) or $\sigma_2 = \sigma_1$. Hence, for each $T = 2, ..., 5$, there are 4 different problems leading to (5-1)*4=16 runs of `MSLP-APPROX`, and the total number L of (independent) random elements varies between 4 and 24 (see third column in Table 6.3). The results are presented here for iteration $k = 0$ and $k = 30$ (for $T = 2$), $k = 20$ (for $T = 3$) and $k = 15$ (for $T = 4, 5$). The other column headers in Table 6.3 denote the same as in the tables of the previous section. The parameters in `Step 0` of `MSLP-APPROX` have been chosen as

$$\varepsilon := 0.01 \cdot (|\mathrm{LB}^{(0)}| + 1),\ N := 200,\ \varrho := 0.05,$$
$$\lambda_{ti} := \begin{cases} 1, \text{ if } i = 1,2,3,4 \\ 2, \text{ if } i = 5,6 \text{ (and } \sigma_2 > 0) \end{cases},\ \forall t \geq 2.$$

Hence, in the refinement steps of the examples where both b_t and B_t are stochastic, the random elements occurring in B_t get a higher weight. Some observations in Table 6.3 are very similar to those of Table 6.1 and 6.2. The smallness of the relative error of the bounds depends strongly on the scale parameters, and the improvements by refining is less effective for larger T. Here again, the increase of the lower bounds is always greater than the decrease of the upper bounds. The first stage candidate solutions at the first and the last iteration for the examples where $\sigma_1 = 2$ are shown in Table 6.4. Note that the variability of some components (cf. $\widehat{x}_{13}$, $\widehat{x}_{14}$ and $\widehat{x}_{15}$) is high relative to the others.

$$A_1 = \begin{pmatrix} -8.999 & & 0.529 & 6.109 & 0.408 & & & 1.952 \\ -4.792 & & 1.034 & 8.999 & & & & -5.241 \\ -4.792 & & 0.779 & 8.999 & 4.773 & -6.998 & -2.761 & \\ 6.154 & 6.099 & & & -5.981 & 3.798 & -10.070 & \end{pmatrix}$$

$$A_2 = \begin{pmatrix} 4.583 & & & 5.439 & -2.016 & & 0.975 & -8.981 \\ 8.999 & 2.448 & -5.323 & 3.498 & 4.100 & -13.723 & & \\ 7.583 & 4.050 & & 8.999 & -9.031 & -11.602 & & \\ 1.766 & & & & 9.296 & -4.604 & & -6.458 \end{pmatrix}$$

$$A_3 = \begin{pmatrix} & & -8.999 & -7.655 & -2.050 & -7.142 & & 25.847 \\ & -1.892 & -3.525 & & & -0.310 & -8.895 & 14.623 \\ & & -8.595 & -8.999 & 17.594 & & & \\ -8.940 & -1.603 & -2.986 & -8.99 & & -0.819 & -2.139 & 25.489 \end{pmatrix}$$

$$A_4 = \begin{pmatrix} & & & 7.396 & 8.870 & 9.787 & -26.055 & \\ -4.566 & & -5.066 & & 5.300 & 2.645 & & 1.686 \\ 4.385 & 8.999 & -5.066 & 2.449 & 7.528 & -6.676 & & -11.620 \\ 7.875 & & -8.999 & -4.289 & 5.414 & & & \end{pmatrix}$$

$$A_5 = \begin{pmatrix} -8.999 & -7.923 & -8.999 & 3.674 & & -6.034 & 28.282 & \\ -5.650 & & -5.650 & & 5.011 & & 6.289 & \\ 3.691 & 4.510 & 3.691 & & 1.557 & & -1.090 & -12.359 \\ 2.547 & & & & 6.438 & 7.200 & -16.186 & \end{pmatrix}$$

$$\mathtt{B}_2 = \begin{pmatrix} -0.599 & 1.934 & & 7.551 & & & 2.886 & 8.999 \\ & -6.986 & & -3.945 & 3.338 & -8.696 & & 3.290 \\ -0.302 & -5.998 & & -1.441 & 4.532 & & 0.025 & 8.999 \\ -1.728 & -5.589 & -1.340 & -5.856 & & & & \end{pmatrix}$$

$$\mathtt{B}_3 = \begin{pmatrix} -8.999 & -8.999 & -9.663 & -0.849 & & & & 4.591 \\ 3.417 & -4.963 & 3.669 & & -4.691 & 4.970 & 6.121 & 9.848 \\ 6.412 & -1.649 & 6.885 & & & 9.328 & 1.739 & 1.777 \\ 5.586 & & & 2.833 & & & & -9.820 \end{pmatrix}$$

$$\mathtt{B}_4 = \begin{pmatrix} & -6.303 & & & 0.250 & & -9.684 & 1.957 \\ & & 8.548 & & 2.814 & -8.999 & -3.018 & 6.914 \\ 4.169 & 3.261 & 5.270 & -5.736 & & 4.004 & -6.847 & \\ & -2.338 & 8.998 & 7.839 & & -4.615 & & 3.545 \end{pmatrix}$$

$$\mathtt{B}_5 = \begin{pmatrix} & & 8.999 & & 3.214 & & 8.999 & 4.351 \\ 1.880 & & 8.874 & 1.145 & 2.337 & 8.999 & 8.874 & \\ & -2.077 & -2.458 & 4.486 & 5.899 & -3.174 & 0.418 & \\ 9.628 & & & 8.603 & & & 5.599 & -4.391 \end{pmatrix}$$

$$\mathtt{b}_1 = \begin{pmatrix} -1.823 \\ 0.730 \\ -3.793 \\ -7.918 \end{pmatrix},\ \mathtt{b}_2 = \begin{pmatrix} -3.200 \\ 5.588 \\ -7.604 \\ 5.647 \end{pmatrix},\ \mathtt{b}_3 = \begin{pmatrix} 4.223 \\ -8.443 \\ 5.037 \\ -8.852 \end{pmatrix},\ \mathtt{b}_4 = \begin{pmatrix} 3.324 \\ -8.462 \\ -7.210 \\ -6.459 \end{pmatrix},\ \mathtt{b}_5 = \begin{pmatrix} -0.700 \\ -2.455 \\ -8.162 \\ -1.544 \end{pmatrix}$$

$$c_1 = \begin{pmatrix} 3.124 \\ 3.023 \\ 1.777 \\ 8.999 \\ 0.734 \\ 0.061 \\ 3.257 \\ 0.091 \end{pmatrix},\ c_2 = \begin{pmatrix} 8.056 \\ 9.558 \\ 4.109 \\ 4.576 \\ 9.731 \\ 1.517 \\ 6.223 \\ 1.073 \end{pmatrix},\ c_3 = \begin{pmatrix} 8.947 \\ 2.476 \\ 9.563 \\ 0.522 \\ 5.181 \\ 1.509 \\ 8.101 \\ 4.061 \end{pmatrix},\ c_4 = \begin{pmatrix} 0.723 \\ 7.584 \\ 0.925 \\ 0.877 \\ 5.504 \\ 2.938 \\ 5.700 \\ 6.358 \end{pmatrix},\ c_5 = \begin{pmatrix} 0.946 \\ 3.340 \\ 1.374 \\ 0.650 \\ 0.652 \\ 8.695 \\ 7.412 \\ 3.329 \end{pmatrix}$$

Figure 6.4: *Data of the test problem.*

Table 6.3: Results for the test problem.

T	stoch.	L	σ_1	k	$\sharp\mathcal{N}_T^{(k)}$	N_k	LB$^{(k)}$	$\overline{\text{UB}}^{(k)}$	$\overline{\Delta}^{(k)}$	$\pm 2{*}\text{SE}^{(k)}$	$\frac{\overline{\Delta}^{(k)}}{\lvert\text{LB}^{(k)}\rvert+1}$
2	b_t ($\sigma_2 = 0$)	4	0.5	0	1	200	24.620	26.258	1.638	0.223	0.064
				30	**748**	**897**	**25.724**	**25.750**	**0.026**	**0.008**	**0.001**
			2	0	1	200	24.620	31.770	7.150	0.873	0.279
				30	**1126**	**897**	**30.304**	**30.392**	**0.088**	**0.025**	**0.003**
	b_t, B_t $\sigma_2 = \sigma_1$	6	0.5	0	1	200	24.620	26.749	2.129	0.319	0.083
				30	**1638**	**897**	**26.231**	**26.375**	**0.144**	**0.029**	**0.005**
			2	0	1	200	24.620	34.430	9.811	1.245	0.383
				30	**2390**	**897**	**31.744**	**32.277**	**0.533**	**0.120**	**0.016**
3	b_t ($\sigma_2 = 0$)	8	0.5	0	1	200	38.043	40.836	2.793	0.288	0.072
				20	**2156**	**544**	**40.198**	**40.460**	**0.263**	**0.065**	**0.006**
			2	0	1	200	38.043	54.482	16.439	1.954	0.421
				20	**3899**	**544**	**50.472**	**52.929**	**2.458**	**0.411**	**0.048**
	b_t, B_t $\sigma_2 = \sigma_1$	12	0.5	0	1	200	38.043	41.353	3.310	0.365	0.085
				20	**2830**	**544**	**39.959**	**40.993**	**1.034**	**0.198**	**0.025**
			2	0	1	200	38.043	58.961	20.918	2.639	0.536
				20	**2506**	**544**	**49.759**	**57.696**	**7.936**	**0.975**	**0.156**
4	b_t ($\sigma_2 = 0$)	12	0.5	0	1	200	41.620	45.155	3.535	0.321	0.083
				15	**971**	**424**	**44.262**	**45.006**	**0.745**	**0.145**	**0.016**
			2	0	1	200	41.620	60.711	19.092	2.169	0.448
				15	**1797**	**424**	**54.733**	**59.447**	**4.714**	**0.736**	**0.085**
	b_t, B_t $\sigma_2 = \sigma_1$	18	0.5	0	1	200	41.620	45.454	3.835	0.400	0.090
				15	**3072**	**424**	**43.929**	**45.433**	**0.240**	**0.240**	**0.033**
			2	0	1	200	41.620	70.409	28.789	3.886	0.675
				15	**4310**	**424**	**54.904**	**67.189**	**12.284**	**1.918**	**0.220**
5	b_t ($\sigma_2 = 0$)	16	0.5	0	1	200	64.291	69.516	5.225	0.569	0.080
				15	**2502**	**424**	**67.932**	**69.438**	**1.506**	**0.261**	**0.022**
			2	0	1	200	64.291	97.590	33.299	4.379	0.510
				15	**4351**	**424**	**84.388**	**93.660**	**9.271**	**1.496**	**0.109**
	b_t, B_t $\sigma_2 = \sigma_1$	24	0.5	0	1	200	64.291	70.442	6.151	0.764	0.094
				15	**2809**	**424**	**65.987**	**70.042**	**4.055**	**0.489**	**0.061**
			2	0	1	200	64.291	115.774	51.483	8.574	0.789
				15	**3787**	**424**	**80.843**	**108.264**	**27.421**	**3.706**	**0.335**

Table 6.4: First stage solution $\widehat{x}_1 \in \mathbb{R}^8$ of the expected value problem (iteration $k = 0$) and at the last iteration for the examples where $\sigma_1 = 2$ and either $\sigma_2 = 0$ (only b_t stochastic) or $\sigma_2 = \sigma_1$.

T	stoch.	$\widehat{x}_{11}$	$\widehat{x}_{12}$	$\widehat{x}_{13}$	$\widehat{x}_{14}$	$\widehat{x}_{15}$	$\widehat{x}_{16}$	$\widehat{x}_{17}$	$\widehat{x}_{18}$
2	-	0.361	0.000	2.378	0.000	0.405	0.464	0.941	0.000
	b_t	0.348	0.000	2.318	0.000	0.198	0.304	0.996	0.000
	b_t, B_t	0.404	0.000	0.186	0.275	0.086	0.270	1.084	0.000
3	-	0.398	0.000	0.544	0.231	0.155	0.315	1.056	0.000
	b_t	0.364	0.000	1.651	0.085	0.140	0.275	1.029	0.000
	b_t, B_t	0.406	0.000	0.038	0.293	0.047	0.244	1.098	0.000
4	-	0.398	0.000	0.544	0.231	0.155	0.315	1.056	0.000
	b_t	0.382	0.000	1.117	0.156	0.169	0.312	1.037	0.000
	b_t, B_t	0.412	0.000	0.000	0.301	0.121	0.302	1.080	0.000
5	-	0.350	0.000	2.326	0.000	0.227	0.327	0.988	0.000
	b_t	0.347	0.000	2.315	0.000	0.187	0.295	0.999	0.000
	b_t, B_t	0.385	0.000	0.856	0.188	0.095	0.261	1.063	0.000

7 Concluding remarks and open questions

For the proposed approximation scheme `MSLP-APPROX` we shall discuss the assumptions M1)-M3) of page 53 and look for possible extensions.

M1)-M2) concern mainly the specific structure of the random data for the purpose of inheritance of dual feasibility by aggregation, see also Theorem 3.14. As we have outlined on page 35, there are even some special situations with randomness in the cost vector, leading to this inheritance after a reformulation of the problem; but the resulting right-hand sides are then not affine linear in the random components. Thus, a possible extension of `MSLP-APPROX` could be to generalize it to the case of non-affine-linear mappings $\big(\mathtt{B}_t(\cdot), \mathtt{b}_t(\cdot)\big)_{t=2,...,T}$. However, in general, it is therefore necessary to discretize the space of the random data $(B_t, b_t)_{t=2,...,T}$ rather than the space of their arguments $(\xi_{ti})_{i=1,...,l_t,\, t=2,...,T}$.

We remark that the validation of the recursive policy $\overline{x}$ of Theorem 4.7 does not rely on the assumption that $(\widehat{u}, \widehat{s}) \in \operatorname{argmax}\big(\mathtt{MSLP}\text{-}\mathcal{D}(\widehat{\mathcal{F}})\big)$ is feasible in the original dual. But the question is how to construct a feasible dual solution in $\mathtt{MSLP}\text{-}\mathcal{D}(\mathcal{F})$ when all data $(B_t, b_t, c_t)_{t=2,...,T}$ are allowed to be random. This might generally be achieved by a backward recursion, similarly to the forward recursion of Theorem 4.7. For this purpose, suppose that $\widehat{x} = (\widehat{x}_1, ..., \widehat{x}_T)$ solves $\mathtt{MSLP}\text{-}\mathcal{P}(\widehat{\mathcal{F}})$ and $(\widehat{u}, \widehat{s}) = \big((\widehat{u}_1, \widehat{s}_1), ..., (\widehat{u}_T, \widehat{s}_T)\big)$ solves $\mathtt{MSLP}\text{-}\mathcal{D}(\widehat{\mathcal{F}})$ where neither of them must be feasible in the original $\mathtt{MSLP}\text{-}\mathcal{P}(\mathcal{F})$ and $\mathtt{MSLP}\text{-}\mathcal{D}(\mathcal{F})$, respectively. Then define $\forall \omega \in \Omega$,

$$\Theta_T(\omega) := \operatorname*{argmin}_{\mathtt{v}_T \in \mathbb{R}^{m_T},\, \mathtt{r}_T \in \mathbb{R}^{m_T}} \left\{ \widehat{x}_T(\omega)^\top \mathtt{r}_T \;\middle|\; A_T^\top \mathtt{v}_T + \mathtt{r}_T = c_T(\omega),\ \mathtt{r}_T \geq 0 \right\},$$
$$\big(\overline{u}_T(\omega), \overline{s}_T(\omega)\big) := \operatorname{argmin}_{(\mathtt{v}_T, \mathtt{r}_T) \in \Theta_T(\omega)} |\mathtt{v}_T - \widehat{u}_T(\omega)|,$$

and for $t = T-1, T-2, ..., 1$,

$$\Theta_t(\omega) := \operatorname*{argmin}_{\mathtt{v}_t \in \mathbb{R}^{m_t},\, \mathtt{r}_t \in \mathbb{R}^{m_t}} \left\{ \widehat{x}_t(\omega)^\top \mathtt{r}_t \middle| A_t^\top \mathtt{v}_t + \mathtt{r}_t = c_t(\omega) - \mathbb{E}[B_{t+1}^\top \overline{u}_{t+1} \,|\, \mathcal{F}_t](\omega), \mathtt{r}_t \geq 0 \right\},$$
$$\big(\overline{u}_t(\omega), \overline{s}_t(\omega)\big) := \operatorname{argmin}_{(\mathtt{v}_t, \mathtt{r}_t) \in \Theta_t(\omega)} |\mathtt{v}_t - \widehat{u}_t(\omega)|.$$

Under appropriate assumptions, the resulting "dual policy" $(\overline{u}, \overline{s}) := \big((\overline{u}_1, \overline{s}_1), ..., (\overline{u}_T, \overline{s}_T)\big)$ turns out to be feasible in the original dual $\mathtt{MSLP}\text{-}\mathcal{D}(\mathcal{F})$. In combination with the primal recursion $\overline{x} = (\overline{x}_1, ..., \overline{x}_T)$ of Theorem 4.7, the duality gap is given by the expectation $\mathbb{E}[\overline{s}^\top \overline{x}]$. But contrary to $\overline{x}(\omega)$, it seems difficult to compute the realizations $\overline{s}(\omega)$ exactly because the recursion requires to evaluate the conditional expectations $\mathbb{E}[B_{t+1}^\top \overline{u}_{t+1} \,|\, \mathcal{F}_t](\omega)$ $(t = 1, ..., T-1)$ which are given by a possibly infinite number of different outcomes $\overline{u}_{t+1}(\omega')$. For this reason, it would probably be convenient to use a conditional sampling procedure to have an estimate $\widetilde{s}(\omega)$ of $\overline{s}(\omega)$. How could this be integrated into the framework of Section 4.2 concerning the tightness of $\widetilde{s}^\top \overline{x} : \Omega \to [0, \infty)$?

In M3) we assume that MSLP-$\mathcal{P}(\mathcal{F})$ has relatively complete (fixed) recourse. In particular, even the slightly stronger version RCR^o is needed according to Definition 4.10. This assumption is rarely verifiable except that $A_t \mathbb{R}_+^{\mathtt{n}_t} = \mathbb{R}^{\mathtt{m}_t}$ $(t = 1, ..., T)$. But the latter would be in contradiction to the boundedness condition (5.1) that is theoretically needed in the case of non-fixed matrices $B_2, ..., B_T$ (see also Remark 4.9). When the RCR^o property fails, at least, one may hope that all linear subproblems (5.13) occurring in MSLP-APPROX turn out to be feasible. Otherwise, it is necessary to add feasibility cuts in terms of an infinite problem, similarly to the finite discrete case by using a nested decomposition scheme (see references in Section 3.1). It is certainly not clear how that can be achieved for continuously distributed data with respect to simulated realizations of decisions. On the other hand, as discussed in Section 4.2, if MSLP-$\mathcal{P}(\mathcal{F})$ is not RCR^o, then one cannot be sure that the aggregated dual solutions $(\widehat{u}, \widehat{s})$ are bounded uniformly with respect to the subfiltrations. In particular, the information for suitable refinements coming from the complementarity variable $\Delta = \widehat{s}^\top \overline{x}$ could be misleading because the worst-case behavior in (5.16) does not need to hold in this case (see also Example 4.14 with the details in Appendix A).

Further research is needed to extend the investigated approximation scheme to the case of nonlinear constraints and/or objectives in MSLP-$\mathcal{P}(\mathcal{F})$. The dual MSLP-$\mathcal{D}(\mathcal{F})$ can then be generalized to the Lagrange dual problem. Theorem 3.14 has been concerned with sufficient conditions for the inheritance of dual feasibility by aggregation; for given independent random vectors $\xi_2, ..., \xi_T$, the result has been that the right-hand side b_t at stage t is allowed to depend on $(\xi_2, ..., \xi_t)$, whereas the technology matrix B_t must not depend on the past $(\xi_2, ..., \xi_{t-1})$. What would be a generalization of Theorem 3.14 to nonlinear constraints?

Another challenge is to deal with integer constraints on some decision variables. An introduction to multistage stochastic integer programming can be found in Römisch/Schultz [28]. It would be interesting to combine a branch-and-bound type algorithm for mixed-integer linear programming with the presented method of aggregation/disaggregation of continuously distributed data. At each iteration of such a scheme one has to decide whether an increase of the number of approximating scenarios is advisable and/or whether some integer variables should be fixed in terms of a branch-and-bound methodology. Hence the goal would be to provide information about the loss due to both aggregating data and relaxing the integrality constraints.

A Details for Example 4.14

We refer to Example 4.14 on page 50, having the problem

$$\texttt{MSLP-}\mathcal{P}(\mathcal{F}): \quad \text{Minimize}_x \ \mathbb{E}\Big[x_1 + x_{2_1}(\xi_2) + x_{2_2}(\xi_2) + x_3(\xi_2, \xi_3)\Big]$$

$$\begin{array}{lcllcl}
x_1 & & & = & 1 & \\
x_1 & + & x_{2_1}(\xi_2) - x_{2_2}(\xi_2) & = & \xi_2 & \text{(a.s.)} \\
 & & x_{2_1}(\xi_2) + x_{2_2}(\xi_2) \quad - \quad x_3(\xi_2, \xi_3) & = & \xi_3 & \text{(a.s.)} \\
 & & x_1, x_2(\xi_2), x_3(\xi_2, \xi_3) & \geq & 0 & \text{(a.s.)} \ ,
\end{array}$$

where ξ_2 and ξ_3 are stochastically independent, ξ_2 takes on the values -1 and 3 with equal probability and ξ_3 is uniformly distributed on $[0,1]$. We have shown in Example 4.14 that $\texttt{MSLP-}\mathcal{P}(\mathcal{F})$ has the RCR^o property according to Definition 4.10, but contrary to Theorem 4.13, Theorem 4.12 is not applicable because $\operatorname{supp}\{b_2\} = \operatorname{supp}\{\xi_2\} = \{-1, 3\}$ is not convex. The dual problem of $\texttt{MSLP-}\mathcal{P}(\mathcal{F})$ is formulated as

$$\texttt{MSLP-}\mathcal{D}(\mathcal{F}): \quad \text{Maximize}_{u,s} \ \mathbb{E}\Big[u_1 + \xi_2 u_2(\xi_2) + \xi_3 u_3(\xi_2, \xi_3)\Big]$$

$$\begin{array}{llllllll}
u_1 & + \ \mathbb{E}[u_2(\xi_2)] & & +s_1 & & & = & 1 \\
 & u_2(\xi_2) & + \ \mathbb{E}\big[u_3(\xi_2, \xi_3) \,\big|\, \xi_2\big] & & +s_{2_1}(\xi_2) & & = & 1 \quad \text{(a.s.)} \\
 & -u_2(\xi_2) & + \ \mathbb{E}\big[u_3(\xi_2, \xi_3) \,\big|\, \xi_2\big] & & +s_{2_2}(\xi_2) & & = & 1 \quad \text{(a.s.)} \\
 & & -u_3(\xi_3, \xi_3) & & & +s_3(\xi_2, \xi_3) & = & 1 \quad \text{(a.s.)} \\
 & & & & & s_1, s_2(\xi_2), s_3(\xi_2, \xi_3) & \geq & 0 \quad \text{(a.s.)}.
\end{array}$$

Both the primal and the dual problem have a unique optimal solution given by

$$\begin{array}{lll lll lll}
\overline{\overline{x}}_1 & = & 1 & \overline{\overline{u}}_1 & = & 1 & \overline{\overline{s}}_1 & = & 0 \\
\overline{\overline{x}}_2 & = & \begin{cases} \begin{pmatrix} 0 \\ 2 \end{pmatrix}, & \text{if } \xi_2 = -1 \\ \begin{pmatrix} 2 \\ 0 \end{pmatrix}, & \text{if } \xi_2 = 3 \end{cases} & \overline{\overline{u}}_2 & = & \begin{cases} -2, & \text{if } \xi_2 = -1 \\ \ \ 2, & \text{if } \xi_2 = 3 \end{cases} & \overline{\overline{s}}_2 & = & \begin{cases} \begin{pmatrix} 4 \\ 0 \end{pmatrix}, & \text{if } \xi_2 = -1 \\ \begin{pmatrix} 0 \\ 4 \end{pmatrix}, & \text{if } \xi_2 = 3 \end{cases} \\
\overline{\overline{x}}_3 & = & 2 - \xi_3 & \overline{\overline{u}}_3 & = & -1 & \overline{\overline{s}}_3 & = & 0 \qquad .
\end{array}$$

A straightforward computation shows that

$$\min\big(\texttt{MSLP-}\mathcal{P}(\mathcal{F})\big) = \mathbb{E}\Big[\overline{\overline{x}}_1 + \overline{\overline{x}}_{2_1} + \overline{\overline{x}}_{2_2} + \overline{\overline{x}}_3\Big] = \frac{9}{2} \ \big(= \max\big(\texttt{MSLP-}\mathcal{D}(\mathcal{F})\big)\big).$$

We now consider an ascending chain $(\widehat{\mathcal{F}}^{(k)})_{k \geq 0}$ of subfiltrations defined by $\widehat{\mathcal{F}}_1^{(k)} = \widehat{\mathcal{F}}_2^{(k)} := \{\emptyset, \Omega\}$ and

$$\widehat{\mathcal{F}}_3^{(k)} := \sigma\bigg(\xi_3^{-1}\Big(\big[\frac{2^{i-1}-1}{2^{i-1}}, \frac{2^i - 1}{2^i}\big)\Big)_{i=1,\dots,k}, \xi_3^{-1}\Big(\big[\frac{2^k - 1}{2^k}, 1\big]\Big)\bigg) \ \ (k = 0, 1, 2, \dots).$$

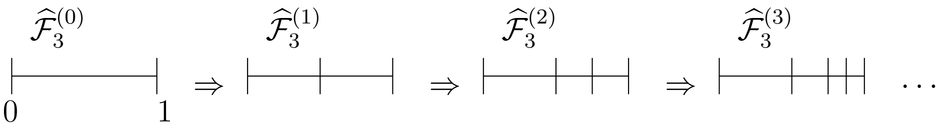

Figure A.1: *Successive splitting of* $[0,1]$ *to which* $(\widehat{\mathcal{F}}_3^{(k)})_{k\geq 0}$ *is characterized.*

The fragmentation of the random support $[0,1]$ of $\xi_3 \sim U[0,1]$ is illustrated in Figure A.1. For $k \in \{0,1,2,..\}$, the probabilities of the $(k+1)$ subintervals of $[0,1]$ are given by

$$\mathbb{P}\Big[\xi_3 \in \big[\tfrac{2^{i-1}-1}{2^{i-1}}, \tfrac{2^i-1}{2^i}\big)\Big] = \tfrac{1}{2^i} \quad (i=1,...,k) \ \text{ and } \mathbb{P}\Big[\xi_3 \in \big[\tfrac{2^k-1}{2^k}, 1\big]\Big] = \tfrac{1}{2^k},$$

and the conditional expectations result in

$$\mathbb{E}\Big[\xi_3 \;\Big|\; \xi_3 \in \big[\frac{2^{i-1}-1}{2^{i-1}}, \frac{2^i-1}{2^i}\big)\Big] = \frac{2^{i+1}-3}{2^{i+1}} \quad (i=1,...,k)$$

and

$$\mathbb{E}\Big[\xi_3 \;\Big|\; \xi_3 \in \big[\frac{2^k-1}{2^k}, 1\big]\Big] = \frac{2^{k+1}-1}{2^{k+1}}.$$

Thus, noting that $\mathbb{E}\xi_2 = 1$, the aggregated problem turns out to be

$$\texttt{MSLP-}\mathcal{P}(\widehat{\mathcal{F}}^{(k)}): \quad \text{Minimize}_x \;\; x_1 + x_{2_1} + x_{2_2} + \textstyle\sum_{i=1}^k \frac{1}{2^i} x_3(i) + \frac{1}{2^k} x_3(k+1)$$

$$\begin{array}{ccccrcll}
x_1 & & & & & = & 1 & \\
x_1 & + & x_{2_1} - x_{2_2} & & & = & 1 & \\
 & & x_{2_1} + x_{2_2} & - & x_3(i) & = & \frac{2^{i+1}-3}{2^{i+1}} & \text{(i=1,...,k)} \\
 & & x_{2_1} + x_{2_2} & - & x_3(k+1) & = & \frac{2^{k+1}-1}{2^{k+1}} & \\
 & & & & x_1, x_2, x_3(i) & \geq & 0 & \text{(i=1,...,k+1)}
\end{array}$$

with its dual formulation

$$\texttt{MSLP-}\mathcal{D}(\widehat{\mathcal{F}}^{(k)}): \quad \text{Maximize}_{u,s} \;\; u_1 + u_2 + \textstyle\sum_{i=1}^k \frac{1}{2^i} \frac{2^{i+1}-3}{2^{i+1}} u_3(i) + \frac{1}{2^k} \frac{2^{k+1}-1}{2^{k+1}} u_3(k+1)$$

$$\begin{array}{rcccrccrcl}
u_1 + & u_2 & & & +s_1 & & & = & 1 & \\
 & u_2 & + & \sum_{i=1}^k \frac{1}{2^i} u_3(i) + \frac{1}{2^k} u_3(k+1) & & +s_{2_1} & & = & 1 & \\
 & -u_2 & + & \sum_{i=1}^k \frac{1}{2^i} u_3(i) + \frac{1}{2^k} u_3(k+1) & & +s_{2_2} & & = & 1 & \\
 & & & -u_3(i) & & & +s_3(i) & = & 1 & (i=1,...,k+1) \\
 & & & & & & s_1, s_2, s_3(i) & \geq & 0 & (i=1,...,k+1).
\end{array}$$

Again, these problems are both uniquely solvable with the solution

$$\begin{aligned}
\widehat{x}_1^{(k)} &= 1 \\
\widehat{x}_2^{(k)} &= \begin{pmatrix} \frac{2^{k+1}-1}{2^{k+2}} \\ \frac{2^{k+1}-1}{2^{k+2}} \end{pmatrix} \\
\widehat{x}_3^{(k)} &= \begin{cases} \frac{3\cdot 2^{k-i}-1}{2^{k+1}} & , \ i = 1, ..., k \\ 0 & , \ i = k+1 \end{cases}
\end{aligned}$$

and

$$\begin{aligned}
\widehat{u}_1^{(k)} &= 1 & \widehat{s}_1^{(k)} &= 0 \\
\widehat{u}_2^{(k)} &= 0 & \widehat{s}_2^{(k)} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
\widehat{u}_3^{(k)} &= \begin{cases} -1 & , \ i = 1, ..., k \\ 2^{k+1}-1 & , \ i = k+1 \end{cases} & \widehat{s}_3^{(k)} &= \begin{cases} 0 & , \ i = 1, ..., k \\ 2^{k+1} & , \ i = k+1 \end{cases} ,
\end{aligned}$$

respectively, leading to the optimal value

$$\min\left(\texttt{MSLP-}\mathcal{P}(\widehat{\mathcal{F}}^{(k)})\right) = \max\left(\texttt{MSLP-}\mathcal{P}(\widehat{\mathcal{F}}^{(k)})\right) = \frac{5}{2} - \frac{1}{2^k}.$$

Since $\widehat{s}_3^{(k)}$ is not uniformly bounded in k, we have a counterexample to Lemma 4.11 noting that assumption B3) is violated. According to (5.13), page 59, and noting that $\widehat{s}_2^{(k)} = (0,0)^\top$, the complementarity part of stage $t = 2$ yields $\Delta_2^{(k)}(\xi_2) = 0$ (a.s.), which results in the solutions of the quadratic problems (5.14) given by

$$\overline{x}_2^{(k)} = \begin{cases} \begin{pmatrix} 0 \\ 2 \end{pmatrix} , & \text{if } \xi_2 = -1 \\ \begin{pmatrix} 2 \\ 0 \end{pmatrix} , & \text{if } \xi_2 = 3 \end{cases} .$$

In both cases, $\xi_2 = -1$ and $\xi_2 = 3$, one has $\overline{x}_{21}^{(k)}(\xi_2) + \overline{x}_{22}^{(k)}(\xi_2) = 2$. Hence, the decision at stage 3 in $\texttt{MSLP-}\mathcal{P}(\mathcal{F})$ is uniquely determined by $\overline{x}_3^{(k)} = 2 - \xi_3$, having

$$\Delta_3^{(k)}(\xi_2, \xi_3) = \widehat{s}_3^{(k)\top} \overline{x}_3^{(k)} = \begin{cases} 0 & , \text{ if } \ 0 \le \xi_3 < \frac{2^k-1}{2^k} \\ 2^{k+1}(2-\xi_3) & , \text{ if } \ \frac{2^k-1}{2^k} \le \xi_3 \le 1 \end{cases} .$$

Thus, for every $k \geq 0$, the recursive policy $\overline{x}^{(k)} = (\widehat{x}_1^{(k)}, \overline{x}_2^{(k)}, \overline{x}_3^{(k)})$ coincide with the optimal solution $\overline{\overline{x}} = (\overline{\overline{x}}_1, \overline{\overline{x}}_2, \overline{\overline{x}}_3)$ of $\texttt{MSLP-}\mathcal{P}(\mathcal{F})$, and the complementarity variable is given by

$$\widehat{s}^{(k)\top}\overline{x}^{(k)} = \Delta^{(k)}(\xi_2, \xi_3) = \Delta_2^{(k)}(\xi_2) + \Delta_3^{(k)}(\xi_2, \xi_3) = \Delta_3^{(k)}(\xi_2, \xi_3) \, .$$

We conclude that the tightness (4.19) of Theorem 4.12 fails because

$$\text{ess sup}\left\{\widehat{s}^{(k)\top}\overline{x}^{(k)}\right\} = \text{ess sup } \Delta_3^{(k)}(\xi_2, \xi_3) \geq 2^{k+1}(2-1) = 2^{k+1} \overset{k\to\infty}{\longrightarrow} +\infty.$$

Bibliography

[1] E.J. Anderson. *Review of Duality Theory for Linear Programming over Topological Vector Spaces*. Journal of Mathematical Analysis and Application, 97: 380-392, 1983.

[2] E.J. Anderson and P. Nash. *Linear Programming in Infinite-Dimensional Spaces*. John Wiley & Sons-Interscience Series in Discrete Mathematics and Optimization, New York, 1987.

[3] H. Bauer. *Wahrscheinlichkeitstheorie*. 4.A., de Gruyter Lehrbuch, Berlin, NewYork 1991.

[4] J.R Birge. *Solution Methods for Stochastic Dynamic Linear Programs*. Systems Optimization Laboratory, Stanford University, Technical Report SOL 80-29 (1980).

[5] J.R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer Series in Operations Research, NewYork, 1997.

[6] A. Brooke, D. Kendrick, A. Meeraus, and R. Raman. *GAMS. A User's Guide*. Technical report, GAMS Development Corporation, Washington DC, USA, 1998. (It can be downloaded from `http://www.gams.com`.)

[7] G.B. Dantzig. *Linear Programming and Extensions*. Princeton University Press, Princeton, NJ, 1963.

[8] O. Fiedler and W. Römisch. *Stability in Multistage Stochastic Programming*. Annals of Operations Research, 56: 79-93, 1995.

[9] K. Frauendorfer. *Barycentric Scenario Trees in Convex Multistage Stochastic Programming*. Mathematical Programming, 75: 277-293, 1996.

[10] P. Fúsek, P. Kall, J. Mayer, S. Sen and S. Siegrist. *Multistage Stochastic Linear Programming: Aggregation, Approximation, and Some Open Problems*. Technical report, Inst. Oper. Res., University of Zurich, 2000. `http://www.unizh.ch/ior`.

[11] H.I. Gassmann. *MSLiP: A Computer Code for the Multistage Stochastic Linear Programming Problem*. Mathematical Programming, 47:407-423, 1990.

[12] H. Heuser. *Funktionalanalysis*. B.G. Teubner, Stuttgart, 1992.

[13] P. Kall. *Stochastic Linear Programming*. Springer-Verlag, Berlin, 1976.

[14] P. Kall and S.W. Wallace. *Stochastic Programming*. John Wiley&Sons, Chichester, 1994. (It can be downloaded from `http://www.unizh.ch/ior`.)

[15] P. Kall and J. Mayer. *SLP-IOR: A Model Management System for Stochastic Linear Programming - System Design*. In A. J. M. Beulens and H.-J. Sebastian, editors, Optimization-Based Computer-Aided Modelling and Design. Lecture Notes in Control and Information Sciences, 174: 139-157, Springer-Verlag, 1992.

[16] P. Kall and J. Mayer. *Stochastic Linear Programming. Models, Theory and Computation*. International Series in Operations Research 80, Springer Verlag, 2005.

[17] D. Klatte and G. Thiere. *Error Bounds for Solutions of Linear Equations and Inequalities*. ZOR - Mathematical Methods of Operations Research, 41: 191-214, 1995

[18] H-S. Lau and A. H-L. Lau. *A semi-analytical solution for a newsboy problem with mid-period replenishment*. Journal of Operational Research Society, 48: 1245-1253, 1997.

[19] O.L. Mangasarian and R.R. Meyer. *Nonlinear Perturbation of Linear Programs*. SIAM J. Control and Optimization, 17: 745-752, 1979.

[20] J. Mayer. *Stochastic Linear Programming Algorithms: A Comparison Based on a Model Management System*. Gordon and Breach Science Publishers, 1998.

[21] P. Olsen. *Multistage Stochastic Programming with Recourse: The Equivalent Deterministic Problem*. SIAM J. Control and Optimization, 14: 495-517, 1976.

[22] P. Olsen. *Multistage Stochastic Programming with Recourse As Mathematical Programming in an L^p Space*. SIAM J. Control and Optimization, 14: 528-537, 1976.

[23] P. Olsen. *Discretizations of Multistage Stochastic Programming Problems*. Mathematical Programming Study, 6: 111-124, 1976.

[24] T. Pennanen. *Epi-Convergent Discretizations of Multistage Stochastic Programs*. Mathematics of Operations Research, 30: 245-256, 2005.

[25] G.Ch. Pflug. *Scenario tree generation for multiperiod financial optimization by optimal discretization*. Mathematical Programming, Ser. B, 89: 251-257, 2001.

[26] R.T. Rockafellar and R.J-B. Wets. *Stochastic Convex Programming: Relatively Complete Recourse and Induced Feasibility*. SIAM J. Control and Optimization, 14: 574-589, 1976.

[27] R.T. Rockafellar and R.J-B. Wets. *Nonanticipativity and L^1-martingales in stochastic optimization probems*. Mathematical Programming Study, 6:170-187, 1976.

[28] W. Römisch and R. Schultz. *Multistage Stochastic Integer Programs: An Introduction*. In M. Grötchel, S.O. Krumke, and J. Rambau, editors, Online Optimization of Large Scale Systems, Springer, Berlin, pp. 581-600. 2001.

[29] A. Shapiro. *Inference of Statistical Bounds for Multistage Stochastic Programming Problems*. Mathematical Methods of Operations Research, 58: 57-68, 2003.

[30] S.E. Wright. *Primal-Dual Aggregation and Disaggregation for Stochastic Linear Programs*. Mathematics of Operations Research, 19: 893-908, 1994.

Curriculum Vitae

Am 6. November 1972 wurde ich, Simon Siegrist, Bürger von Fahrwangen AG, als Sohn von Samuel und Liselotte Siegrist, geb. Haller, in Aarau AG geboren.

Ausbildung/Tätigkeit

1979-1984 Primarschule in Aarau

1984-1988 Bezirksschule Aarau

1988-1992 Alte Kantonsschule Aarau (Typus E)

1992-1998 Universität Zürich

- Hauptfach: Mathematik
- Nebenfächer: Operations Research und Volkswirtschaftslehre
- Diplomarbeit in *Computational Biology* zum Thema *"Ein symmetrischer Test auf Verwandtschaft zwischen Familien von Sequenzen"* unter der Leitung von Prof. Dr. A.D. Barbour

1999-2004 Assistent am Institut für Operations Research der Universität Zürich

2002-2005 Doktorand an der Universität Zürich

Die vorliegende Dissertation zum Titel *"A Complementarity Approach to Multistage Stochastic Linear Programs"* entstand unter der Leitung von Prof. Dr. A. D. Barbour und Prof. Dr. P. Kall.

Während des Diplom- und Doktorandenstudiums habe ich im mathematischen Bereich Vorlesungen von folgenden Dozenten besucht:

In Mathematik:
A.D. Barbour, P. Gabriel, M. Hieber, H. Jarchow, V. Schroeder, H.H. Storrer;

In Operations Research:
P. Kall, D. Klatte, J. Mayer.

E-Mail: simonsiegrist@bluewin.ch